The Dissertation Committee for Qiu Wu
certifies that this is the approved version of the following dissertation:

Computational Image Analysis of Mass Lesions on Dynamic Contrast-Enhanced Breast MRI

Committee:

Donald S. Fussell, Supervisor

Gary J. Whitman, Supervisor

Alan C. Bovik

Omar Ghattas

Mia K. Markey

Tom E. Milner

Computational Image Analysis of Mass Lesions on Dynamic Contrast-Enhanced Breast MRI

by

Qiu Wu, B.S., M.S.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

December 2009

Acknowledgments

This dissertation is a report of journey on which I have embarked since summer 2004. During this period, I was fortunate enough to meet and work with a collection of remarkable people, most especially, Donald Fussell, Mia Markey, and Gary Whitman. Words can't express my gratitude to my advisor, Prof. Fussell, whose knowledge, insight, vision, and support has guided me through my dissertation period. I benefitted very much from working with Prof. Markey, her expertise on breast image analysis and her advice significantly improved my dissertation work. Due to her emphasis on maintaining clinical focus on the research and her connection with M. D. Anderson Cancer Center (MDACC), I was sent to be a trainee at MDACC to work with an excellent radiologist and devoted mentor, Dr. Whitman. Additionally, I'd like to thank Omar Ghattas, who had a great deal to do with my intellectual development and has a lot impact on me by opening my eyes toward computational and applied mathematics. I appreciate Dr. Bovik for serving on my committee, and I benefited from his high intellectual standard, and excellent teaching in the classroom. I also thank Prof. Milner for his time, I'm impressed by his affection to his students while working as his TA in Spring 2006. In summary, each and every of my committee members is and will be my role model in many aspects.

My first attempt on breast MR image analysis began in the summer of 2004, then I had a fortuitous internship with Computer-Aided Diagnosis Group of Siemens Medical Solutions USA (Malvern, PA). I thank the support from my then manager Marcos Salganicoff, and department head Arun Krishnan. I also appreciate many helps and friendship from then colleagues, such as Jonathan Stockel, Congjin Chen and Jinbo Bi.

Most of dissertation research was conducted when I based at MDACC in Houston. I was exposed to broad medical imaging and breast cancer related education and research; in particular I want to thank Drs. Middletown, Stephens, Dogan, Ma, and Stafford.

None of this work would have been possible without the assistance of a great number of people. I received a lot of encouragement and support from administrators at department, engineering school, and university, most especially, from Prof. Dean Neikirk, the graduate advisor and GSC chair of ECE department. Many staff members at both UT Austin and M. D. Anderson provided valuable administrative support, particularly I want to thank Melaine Gulick of ECE department, Georgia Hartman and Shannon Kawa of Internation Office of UT, Barbara Mahinda of breast imaging section of MDACC. Although mere thanks are inadequate, I thank them.

Many friends kindly provided tremendous help to me in the past years, partic-

ularly I need to acknowledge Prof. Jack Xin (now at UC Irvine), Zhou Wang, Kui Ren, Shulin Ni, Fengfeng Tu. During the past year I had a lot of happy moments at Rice University, and developed my basketball skill from zero to ϵ level in Rice gym. I want to thank Guo-Ping Chang and Tingting Chang for being close friends during the past year. I also want to acknowledge many valuable helps from Chia Chu and her family.

Last but most, I want to thank my parents and my older brother for their endless love and support. I'm certain that they are as happy as I am in my finishing the degree. This dissertation is a dedication to them.

Computational Image Analysis of Mass Lesions on Dynamic Contrast-Enhanced Breast MRI

Publication No. _____

Qiu Wu, Ph.D.
The University of Texas at Austin, 2009

Supervisors: Donald S. Fussell
Gary J. Whitman

This dissertation presents results of a medical image analysis project leading towards development of a comprehensive set of methods and tools for computational image analysis of dynamic contrast-enhanced (DCE) breast magnetic resonance image (MRI), with the aim to aid the physician in interpreting DCE breast MRI examinations. Toward this goal, we developed image analysis methods that would be needed in a breast MRI computer aided diagnosis (CADx) system. A novel contribution of this dissertation is the performance evaluation for each of the major algorithm components developed in this dissertation project.

This dissertation begins with reviewing breast imaging techniques, including routinely used modalities in current clinical practice and emerging techniques still in development. We discuss at length the principles of DCE breast MRI,

a very sensitive breast imaging modality that has been increasingly used in clinical practice. Then we review the diagnostic guidelines for interpreting DCE breast MRI, and explain the needs and challenges that arise in developing computational image analysis system for breast MRI applications.

In this dissertation project, both the morphological and kinetic features of the lesion are automatically extracted for diagnostic purpose. In order to extract morphological features from the segmented lesions, the lesion needs to be accurately segmented out from its surrounding tissues. We utilized a probabilistic method to obtain an optimal segmentation map based on several algorithmic segmentation outputs. In evaluating the performance of segmentation algorithms, we compared the algorithmic segmentation results against manually segmented lesions, and further assessed the segmentation impact on subsequent classification stage. In order to extract accurate kinetic information, the motion needs to be compensated across image volumes acquired sequentially. In this dissertation, we comparatively assessed the similarity metric in registering DCE breast MR images. The performance of cross correlation(CC) coefficient, and mutual information (MI) were studied in both rigid and non-rigid registration schemes. Numerical results and statistical properties were reported. The resultant image quality after registration is discussed both qualitatively and quantitatively. In this dissertation we implemented a classification system based upon quantitative morphological and kinetic features in improving the specificity of breast MRI. Morphological and kinetic features of the lesion

were extracted automatically, and then the feature selection step was utilized to select the most relevant features to maximize the classifier performance. In our study, the area under the receiver operating curve (AUC) is used as the performance metric of the classifier, and our results are competitive with those of previous studies.

The dissertation concludes by summarizing the contribution of this project and suggesting the future directions of quantitative and highly automated approaches to breast MR image analysis.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation and Research Goals

Breast magnetic resonance imaging (MRI) is the most sensitive imaging modality for diagnosing primary and recurrent breast cancer. Breast MRI outperforms other screening modalities for women at high risk of developing breast cancer [64, 69]. Breast MRI has also evolved as a useful tool in staging of breast cancer. Previous studies indicate that breast MRI is more accurate than mammography and ultrasound in estimating tumor size [7, 33, 34, 56, 80]. Thus, breast MRI has promise for improving breast cancer management.

Breast MRI is performed with dynamic contrast-enhanced techniques, in which the breast is scanned before, and after the administration of a contrast agent. Dynamic contrast-enhanced (DCE) breast MRI provides a noninvasive assessment of microcirculatory characteristics of tissues in addition to anatomical information [113]. Traditional manual interpretation of DCE breast MRI is time-consuming and tedious and can lead to oversight error due to the large size of the four-dimensional data sets (three spatial dimensions plus time). Manual interpretation of lesion morphological features and size is also subject

to inter- and intra-observer variability [57, 99, 108]. Currently, the specificity of breast MRI depends on the experience of the radiologist [59].

To facilitate the use of MRI for improving breast cancer care, there is a need to develop an image analysis system to help interpret the images in a more accurate, efficient, and consistent way. Another benefit of such a system would be to reduce the high cost associated with breast MRI exams. Thus, the specific aims of this study are to:

(1) Develop an efficient lesion segmentation method to segment out the lesion from the surrounding tissues on the first post-contrast breast MR image. This step will lead to accurate delineation of the lesion such that morphological or margin features of lesion can be extracted for subsequent lesion classification.

(2) Comparatively study the similarity metric in registering DCE breast MR images to deal with intensity changes over time and motion effects across image volumes acquired sequentially, accuracy and robustness of registration performance will be assessed.

(3) Construct models to classify lesions as benign or malignant, based upon the morphological parameters extracted from the segmented lesion and the kinetic features extracted from the registered breast MR image series.

This study is significant because breast cancer is a major public health problem and the efficiency and accuracy of breast MRI interpretation need to be improved. If successful, this study will enable the development of clinical

decision support systems for breast cancer diagnosis and management.

1.2 Outline of the Dissertation

The main body of this dissertation contains five chapters, chapter 2–6. Chapter 2 reviews the breast imaging techniques, including commonly used ones in current clinical practice and emerging techniques still in development. Chapter 2 discusses extensively the principle of the dynamic contrast-enhanced (DCE) breast MRI. Chapter 3 begins by reviewing the diagnostic guidelines for interpreting breast MRI, and then discusses the needs and challenges that arise in developing computational image analysis system for breast MRI application. In chapter 3, the research goals and design methodologies of this dissertation are outlined. In order to extract morphological features from the segmented lesions, the lesion needs to be accurately segmented out from its surrounding tissues. Chapter 4 utilized a probabilistic method to obtain an optimal segmentation map based on several algorithmic segmentation outputs. In evaluating performance of segmentation algorithms, we compared the algorithmic segmentation results against manually segmented lesions, and further assess the segmentation impact on subsequent classification stage. In order to extract accurate kinetic features on DCE breast MRI, image registration is needed to compensate motion across image volumes acquired sequentially. In chapter 5, we comparatively assessed the similarity metrics in registering DCE breast MR images. The performance of cross correlation(CC) coefficient, and mutual information (MI) were studied in both rigid and non-rigid registra-

tion schemes. Numerical results and statistical properties are reported. The resultant image quality after registration is discussed both qualitatively and quantitatively. Chapter 6 presents a classification system based upon quantitative morphological and kinetic features in improving the specificity of breast MRI. Both morphological and kinetic features of the lesion are extracted automatically, and then a feature selection step was utilized to select the most relevant features to maximize the classifier performance. In our study, the area under the receiver operating characteristic curve (AUC) is used as the performance metric of the classifier, and our results are competitive with those of previous studies. Chapter 7 summarizes the contribution of this dissertation and suggests future directions regarding breast MRI analysis.

Chapter 2

Breast Cancer and Breast Imaging

Breast cancer is the second leading cause of cancer deaths in American women today. Approximately 200,000 new breast cancer cases are diagnosed and about 40,000 women die from breast cancer in the United States. each year [98]. The key to increasing survival is early detection and treatment of breast cancer. Medical imaging is essential to breast cancer screening, diagnosis, and treatment. Several imaging modalities, such as X-ray mammography, ultrasound, and MRI, are used for detection and diagnosis of breast cancer; however, none of the modalities is flawless. An ideal breast imaging modality should detect, localize, and characterize abnormal tissues in the breast. Current technologies cannot meet all three of these requirements in a single imaging modality, so developers typically specialize in optimizing one goal at a time [77]. Thus far, x-ray mammography is the first-line modality for breast cancer screening. Other modalities, including ultrasound and MRI, are used as adjunctive screening techniques [78]. In this chapter, we begin by briefly discussing breast anatomy and breast cancer, and then discuss several commonly used breast imaging modalities, including X-ray mammography, DCE breast MRI, and breast ultrasound. We also discuss three emerging breast imaging techniques: digital breast tomosynthesis, breast CT, and breast PET.

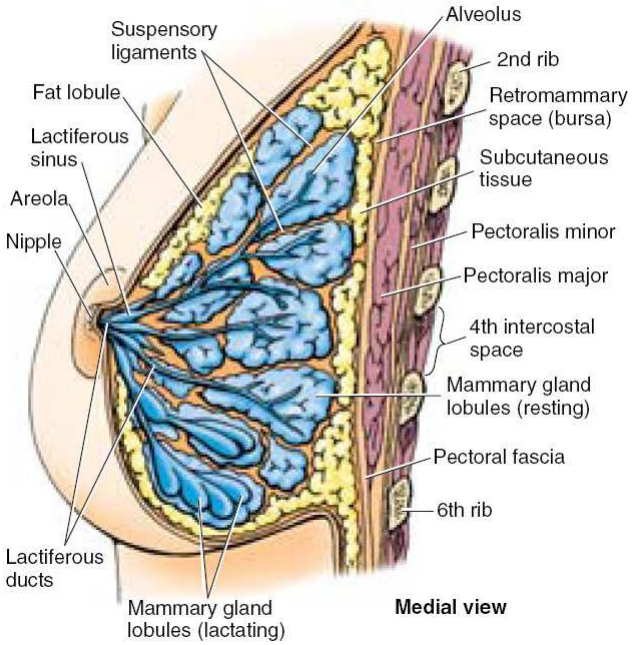


Figure 2.1: Sagittal section of a female breast. Picture courtesy of [73]

2.1 Breast Anatomy and Breast Cancer

2.1.1 Breast Anatomy

The breast is a mass of glandular, fatty, and fibrous tissues positioned on the chest wall extending from the second to the seventh ribs. The breast consists of glandular tissue, fibrous tissue, adipose tissue, lobes, blood vessels, lymphatic vessels, and nerves. Fig. 2.1 illustrates a sagittal section of a female breast and the anterior thoracic wall.

The breast has no muscle tissue. Immediately beneath the skin is a layer of subcutaneous fat, and the fat (adipose tissue) surrounds the glands and determines the shape and the size of the breast. Ducts and lobules comprise the glandular tissue. There are about 15 lobes arranged in a circular pattern in each breast, and each lobe consists of many lobules. The milk is produced by the lobules and is transported by the ducts from the lobules to the nipple. There are between 6 to over 20 duct openings around the nipple [58]. Fibrous tissues are connective tissues surrounding the lobules and the ducts.

Regarding the vascular supply of the breast, the outer half of the breast is supplied by the axillary artery extending from the armpit; the central and the medial portions of the breast are supplied by the internal mammary artery and its perforating branches. Most lymphatic channels in the breasts drain to the axillary lymph nodes, and most lymph drainage passes through one or two lymph nodes, the so called sentinel nodes [58].

2.1.2 Breast Cancer

Cancer derives from a collection of multiple genetic aberrations. It is believed that breast cancer is a genetic disease and is caused by deoxyribonucleic acids (DNA) changes. Such DNA changes can be inherited from family members or acquired during one's life. So far, the exact causes of acquired DNA mutations remain unclear despite years of effort in studying breast cancer. However, some risk factors have been identified including gender, age, genetic risk fac-

tors, family history, race, dense breast tissue, and menstrual history. Some risk factors relate with one's lifestyle, such as having no children or having children after a 30 years, use of birth control pills, hormone therapy, no breast-feeding, alcohol usage, and obesity.

Breast lesions can be generally categorized by the major components in which they develop [58]. Duct ectasia, cystic dilatations of the large duct, large duct papillomas, and intraductal carcinomas are lesions of the major ducts. Hyperplasia, peripheral duct papillomas, and invasive ductal carcinoma are lesions of the minor and the terminal ducts. Cysts, fibroadenomas, adenosis, phyllodes tumors, and lobular carcinoma are the lesions of the lobule. Sarcomas arise from the interlobular connective tissue.

There are several commonly used terms in breast cancer management to describe breast cancer, such as carcinoma, adenocarcinoma, carcinoma in situ, invasive (infiltrating) carcinoma, and sarcoma. Carcinoma is used to describe a cancer that begins in the lining layer of organs, such as the breast. Adenocarcinoma refers to the cancer that begins within the glandular tissue, including the ducts and the lobules. Carcinoma in situ indicates the process is still in the layer of cells where it started, including ductal carcinoma in situ (DCIS) and lobular carcinoma in situ (LCIS). Invasive (infiltrating) carcinoma refers to those cancers growing beyond the layer of cells where it began. Sarcomas are cancers starting from connective tissues.

Typically, hyperplasia, sclerosing adenosis, apocrine metaplasia, fibroadenomas, intraductal papillomas, stromal fibrosis, radial scar, inflammatory changes, atrophic changes, intramammary lymph nodes, and some phyllodes are benign lesions. Invasive ductal carcinoma, DCIS, tubular carcinoma, invasive lobular carcinoma, and medullary carcinoma are malignant lesions. Breast cancer typically spreads through lymphatic vessels, which transport cancer cells from the breast to the lymph nodes, and then to distant locations in the body. The tumor cells then colonize at the secondary sites, producing nests of tumor cells. This process of tumor cells spreading is called metastasis, which is the main process that leads to mortality in breast cancer patients.

The key to breast cancer survival is early detection and treatment. Breast imaging techniques are critical for the early detection of breast cancer. Currently x-ray mammogram is the primary screening modality.

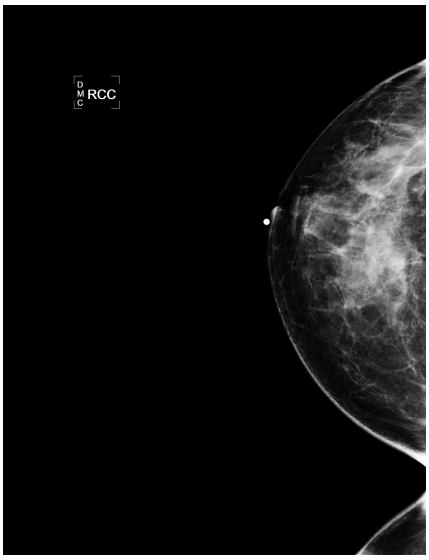
2.2 Mammography

Mammography is a radiographic examination technique that uses X-rays to image breast tissue. The mammographic imaging unit consists of a x-ray tube, a collimator, a compression system, and a detector. The attenuation of the x-ray differs among various breast tissues, providing contrast on mammography [58]. Denser tissue absorbs more X-rays and results in a bright region on the image. Mammography attempts to identify the structural or morphological

differences that can indicate the presence of cancer, such as masses, microcalcifications, and architectural distortions [16].

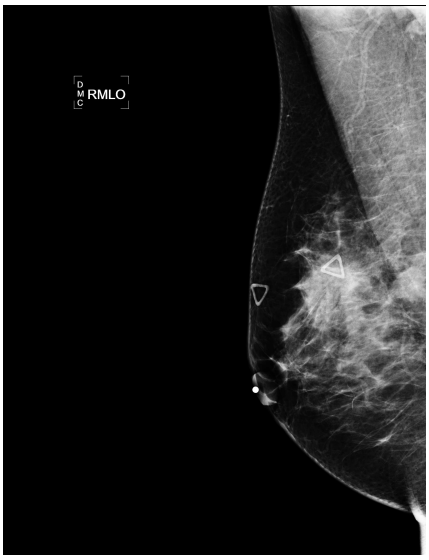
Mammography is a well-developed technology that offers high-quality images at low radiation doses. Mammography can be used for both diagnostic and screening applications. Screening mammography is used to detect abnormalities in woman who are asymptomatic with no signs or symptoms to suggest breast cancer. Screening generally includes two views of each breast, the mediolateral oblique(MLO) projection and the cranio-caudal (CC) projection. Additional views, such as mediolateral (ML), inferosuperior oblique (ISO), from below (FB), lateromedial oblique (LMO), and lateromedial (LM) views, may provide further diagnostic value when abnormality is suspected on the screening mammogram. Early detection by mammographic screening has been shown to result in reduced mortality from breast cancer.

There are two major inherent limitations of x-ray mammography: the information loss of visualizing a 3D structure in 2D projections and the lack of functional insight regarding the biological processes of the breast tissue imaged. About 10-30% of tumors cannot not be detected by mammography [8, 46, 54, 90]. It is particularly difficult to detect tumors in dense breasts and breasts with implants and to identify residual and recurrent cancer using mammography [31, 89]. It is also difficult to differentiate benign from malignant lesions on mammography. The positive predictive value of mammography is



(a) Right CC

(b) Left CC



(c)Right MLO

(d)Left MLO

Figure 2.2: CC and MLO views of a patient diagnosed with invasive ductal carcinoma with DCIS in right breast.

low; only 10-30% of mammographically detected lesions subjected to biopsy prove to be malignant [85]. Tumor size may also be inaccurately estimated on mammography [7, 31].

Adjunctive modalities, including breast ultrasound and breast MRI, can overcome the limitations of mammography to some extent.

2.3 Breast Ultrasound

Ultrasound is a well-established medical imaging technology that uses sound waves to create images of tissue boundaries by detecting reflected sound waves from internal tissues. When sonography is performed by emitting ultrasound pulses into the target tissue, the sound wave reflects at the boundaries between tissues with different acoustic properties. Since the depth of these boundaries is proportional to the time intervals of reflection arrivals, ultrasound can be used for imaging the tissue boundaries [77].

Ultrasound is excellent for differentiating solid masses from cysts, and it is routinely used for cyst/solid differentiation. Also ultrasound is routinely used for guiding interventional procedures, such as fine needle aspiration and core needle biopsy [58].

Ultrasound does not use ionizing radiation, and it is safe and relatively pain-

less. The United States Food and Drug Administration (FDA) has approved ultrasound imaging as an adjunct to mammography for the detection of breast cancer; so far, ultrasound is the most widely used adjunctive modality for breast imaging.

2.4 Breast MRI

2.4.1 MRI

The fundamental principle of magnetic resonance imaging is the nuclear magnetic resonance (NMR) phenomena [52]. The charged and spinning particles, such as protons, have a magnetic dipole moment. When placed in an external magnetic field, these charged and spinning particles rotate, in so called Larmor frequency, along an axis either parallel or anti-parallel to the direction of external magnetic field. The energy level of those in parallel direction is slightly lower than that of anti-parallel direction. I.I. Rabi, in the late 1930s, was the first to apply radiofrequency (RF) pulse of just Larmor frequency to flip nuclear magnetic dipole moments from one orientation to another orientation, this phenomena is so called "nuclear magnetic resonance." There are slight more low-energy state spins than high-energy state ones, therefore tissue net magnetization was induced when human body is placed in an external magnetic field. An RF pulse at Larmor frequency can flip the magnetization away from the longitudinal direction (direction of external field) into the transversal plane (perpendicular to external field). Then the longitudinal magnetization

recovers with a time parameter $T1$ (the longitudinal relaxation time); and the transverse magnetization decrease with a parameter $T2$ (the transverse relaxation time). Usually T1 and T2 are unique biophysical characteristics of the tissue and thus can be used to provide contrast between different tissues. In the 1970s, pioneering contributions have been made leading to the applications of magnetic resonance in medical imaging. Of particular note were two researchers, Paul Lauterbur and Peter Mansfield, who made fundamental contributions. Paul Lauterbur introduced spatially-encoded magnetic field and then resolved the source of the signal. Peter Mansfield showed how the radio signals from the MRI can be mathematically analyzed as a useful image, and he also developed a fast imaging technique called echo-planar imaging (EPI). In 2003, Paul C. Lauterbur and Peter Mansfield shared the Noble Prize in Medicine for their discoveries concerning MRI.

A basic MRI system consists of a static magnetic field, gradient fields (coils), RF coils(transceiver), amplifiers to drive the gradient and RF coils, and a computer to synchronize the pulse sequence and to process the signals. The static magnetic field is a relatively constant and stable magnetic field. To spatially encode the signal, gradient coils are used to modulate the magnetic field strength as a function of space and time. MRI primarily images the NMR signals from the hydrogen nuclei of the tissue. By applying a 3D encoded magnetic field, MRI enables a truly 3D examination of the breast tissues. However, when first imaging breast using MRI technique twenty years ago, the relax-

ation times of normal breast tissue and cancer were not sufficiently different [58], therefor breast MRI without intravenous contrast material does not provide sufficient sensitivity or specificity; and non-contrast enhanced breast MRI alone is not recommended in evaluating for breast cancer (Fig. 2.3).

2.4.2 Angiogenesis and Dynamic Contrast-Enhanced Breast MRI

The need to image tissue function is the driving force behind the development of DCE MRI. Functional imaging is required in order to recognize the distinctions between tumors and normal tissue that exist at the molecular level, such as differences in cellular composition, permeability, and microvessel density. The use of contrast agents in MRI enables the visualization of functional changes, particularly angiogenesis, when dynamic MR images are acquired.

Angiogenesis is a sequential processes beginning with the tumor cells and host cells releasing angiogenic molecules, such as vascular endothelial growth factor (VEGF), to stimulate the endothelial cells to form new small vessels [115]. Tumor cells rely on these capillary sprouts to bring nutrition and oxygen supplies, and carry away the waste materials. Usually, benign neoplasms are sparsely vascular and slow-growing, while malignant neoplasms are highly vascular and fast-growing. Consequently, tumor microcirculation differs profoundly from that of normal organs in three ways: a) in the flow characteristics and sometimes the blood volume of the microvasculature; b) in microvascular permeability; and c) for many malignant tumors, in the increased fractional volume

Figure 2.3: Axial, coronal, and sagittal views of a non-contrast enhanced T1-weighted pre-contrast 3D breast MRI

of extravascular extracellular space (EES).

In a DCE MRI exam, a contrast agent such as gadolinium diethyltriamine-pentaacetic acid (Gd-DTPA) diffuses into the extravascular extracellular space (EES) via the capillaries, accumulates in the tissues with high vascularity, subsequently leaks back into the vascular space, and is eventually excreted from the body (Fig. 2.4). The diffusion process is governed by the kinetic properties of the target tissues. The concentration of the contrast agent alters the relaxation time of the water protons in the surrounding tissue; thus, the accumulated amount of the contrast agent around the targeted tissue is reflected in the signal intensity. As there is contrast agent uptake and washout over time, dynamic MR images are acquired along the time. In DCE MRI, pulse

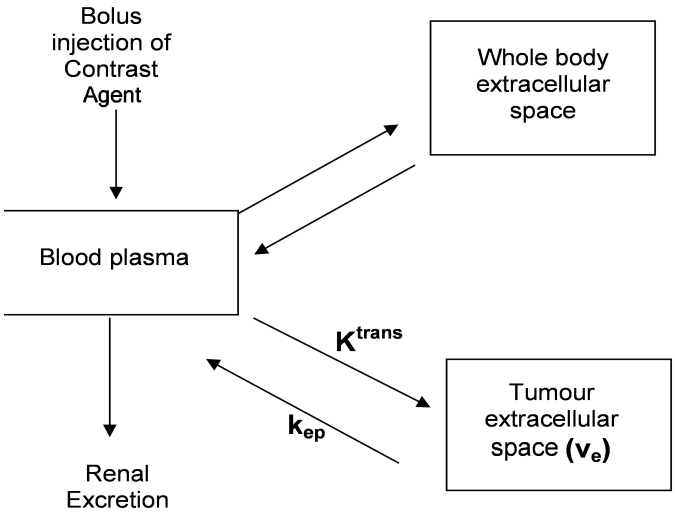


Figure 2.4: Illustration of a compartment model demonstrating exchange of the contrast agent between the blood plasma and the extravascular extracellular space (EES). The contrast agent exchanges from the plasma to the EES at rate K^{trans} and from EES to plasma at rate k_{ep}. The relationship between those two rates is governed by: $k_{ep} = \frac{K^{trans}}{v_e}$ where v_e is the fraction of tumor space occupied by the EES.

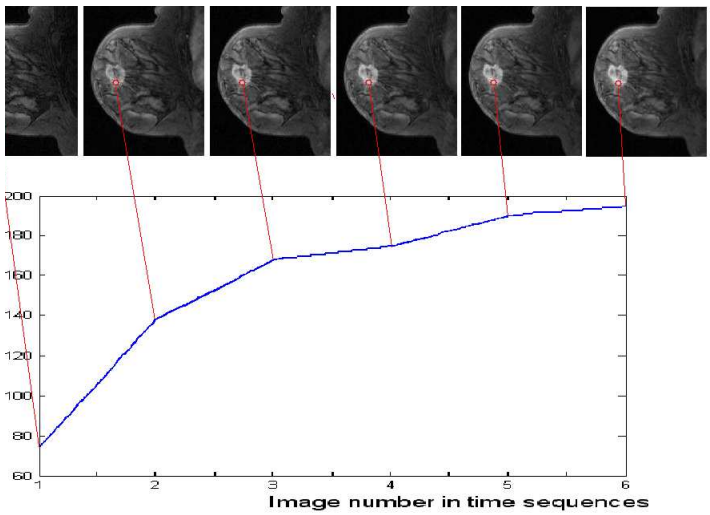


Figure 2.5: An image intensity curve across time for the voxel centered by the red circles. The horizontal axis represents time while the vertical axis represents signal intensity.

sequences can be chosen so that the resulting images are selectively sensitive to different vascular and kinetic characteristics. With appropriate use of contrast agents, dynamic T1- and T2-weighted images can be acquired. The net effects on the post-contrast image intensities of targeted tissues in the T1-weighted and T2-weighted images are opposite [25, 72]. On T1-weighted images there is "enhancement", while on T2-weighted images there is "darkening" [25, 72]. However, the term dynamic contrast-enhancement is used in either case. T1-weighted images have been most commonly used in assessing breast tissue. Unless otherwise stated, the DCE breast MR images discussed in this dissertation are T1-weighted images. Essentially, the dynamic image intensities reflect the physiological nature of the targeted tissue. For any location in the image, the dynamic signal intensities can be viewed as a multi-dimensional signal, usually referred to as the *kinetic enhancement curve* (Fig. 2.5). Previous studies have shown that the kinetic enhancement curves of malignant and benign lesions are usually different. Thus, in addition to the morphological features of the lesion, kinetic features are recommended for use in determining the pathological nature of the lesion.

2.4.3 MRI Contrast Agent and Safety Concern

In 2006, a rare disease, known as nephrogenic systemic fibrosis or nephrogenic fibrosing dermopathy (NSF/NFD), was reported to occur in patients with moderate to end-stage kidney disease after they had a MRI or magnetic resonance angiography (MRA) examination with a gadolinium-based contrast

agent. Consequently, new governmental guidelines have been issued, and existing MRI procedures have been revised on using contrast agents in MRI. The International Society of Magnetic Resonance in Medicine maintains a webpage with a collection of materials and updated information on NSF and MRI safety, to which we recommend interested readers for detailed and updated information on this topic [1].

Though it is not determined yet on the causative relationship between NSF/NFD and gadolinium-based contrast agents, the US FDA [2] has requested the manufacturers to include a new boxed warning on the product labeling of all gadolinium-based contrast agents that are used for enhancing MR images. The requested warning states that "the patients with severe kidney insufficiency who receive gadolinium-based agents are at risk for developing a debilitating, and potentially fatal disease known as NSF." The warning also states that "patients just before or just after liver transplantation, or those with chronic liver disease, are also at risk for developing NSF if they are experiencing kidney insufficiency of any severity."

The FDA release [2] further states that patients should be screened for kidney problems prior to the administration of a gadolinium-based contrast agents. The implementation of appropriate procedures is vital to avoid and prevent accident/incidents. The FDA release [2] requests that the the contrast agent dose not exceed the recommended dose and that enough time elapse be en-

Contrast Agent	Marketed as	Manufacturer
Gadobenate Dimeglumine	MultiHance	Bracco Diagnostic Inc., Princeton, NJ
Gadodiamide	Omniscan	GE Healthcare, Chalfont St. Giles, U.K.
Gadopentetate Dimeglumine	Magnevist	Bayer Schering Pharma, Berlin, Germany
Gadoteridol	ProHance	Bracco Diagnostic Inc., Princeton, NJ
Gadoversetamide	OptiMARK	Mallinckrodt, Inc., Haelwood, Mo.

Table 2.1: Five gadolinium chelates approved by US FDA for use in MRI scan

sured before injection of the contrast agent again.

In the United States, there are five FDA-approved gadolinium-containing contrast agents for use with MRI. Some clinical data suggest that the linear and the nonionic agents, such as Gadodiamide, are more likely to be associated with NSF [10]. Currently, it is not yet clearly determined whether gadolinium-based contrast agents can cause NSF. However, precautions should be taken when using the five gadolinium-based contrast agents in patients [96].

At The University of Texas M. D. Anderson Cancer Center (UTMDACC) , the practice of administering gadolinium-based contrast agents has been revised according to increased awareness regarding NSF. At UTMDACC, patients are questioned whether they have history of renal disease/failure/dialysis or acute renal insufficiency of any severity due to the hepato-renal syndrome or if they are in the perioperative liver transplantation period. Those with no or minimal risk factors can proceed with standard MRI protocols. Those at risk for developing NSF need to be carefully checked before MRI scan with the

administration of a gadolinium-based contrast agent. The glomerular filtration rate (GFR) is an indication of renal function and used in determining the chronic kidney stage(CKD). A GFR value greater than $60ml/min/1.73m^2$ will be considered as normal GFR or Stage 1-2 CKD; these patients can proceed with standard MR scans but not be injected with Omniscan. Those with GFRs between 30 and $59ml/min/1.73m^2$ are considered as Stage 3 CKD; these patients should be reviewed by radiologist individually and no Omniscan administered, but low risk contrast agents, such as Magnevist, or ProHance, or MultiHance, can be considered. Those with GFR less than $29ml/min/1.73m^2$ are considered as Stage 4-5 CKD; these patients need to be carefully reviewed according to the following procedures: the referring or radiology physician determines the Gd-based MRI scan is medically necessary, then a renal physician or nephrologist should be consulted and a radiologist will consider the results of this consult in assessing the risks and the benefits to the patients. Also only low risk gadolinium-based contrast agents can be used and followed by hemodialysis within 2 hours of completion of MR study for the patients at Stage 4-5 CKD.

In additional to breast MRI, other imaging modalities, such as digital breast tomosynthesis and breast computed tomography (CT), can also obtain 3D images of the breast.

2.5 Digital Breast Tomosynthesis and Dedicated Breast CT

Digital breast tomosynthesis (DBT) and Breast CT are two new breast imaging modalities. The setup of DBT is essentially a modification of the conventional mammogram machine [83]. In a DBT scan, the breast is positioned the same way it is in a conventional mammogram, and the X-ray tube moves in an arc around the breast. Multiple X-ray images of each breast from different angles are acquired and then sent to a computer to reconstruct a volumetric image. The total radiation dose for a DBT scan is slightly higher than that for conventional mammogram. The prototype DBT machines have been made by manufacturers for clinical evaluations. These DBT units have dual functions of performing both digital mammograms and breast tomosynthesis with the same machine.

Very recently, dedicated breast computed tomography (CT) systems capable of cone-beam CT of the breast have been developed [11, 116]. In cone-beam breast CT setup, the patient lies prone and breasts are put into the openings of the bed, both x-ray source and flat panel detector system simultaneously rotate 360° around the breast. In cone-beam CT scan, only breasts expose to radiation and the total radiation dose comparable to conventional mammogram [11]. Cone-beam breast CT can provide a true 3D image with isotropic resolution. Total 500 projections are acquired and then sent to computer to reconstruct a 3D image. In the initial breast CT clinical trial, the breast CT

does better in visualizing masses compared to mammography, but has challenge in imaging microcalcifications [66].

Both digital breast tomosynthesis and breast CT are promising techniques and are still under development; larger clinical trials are needed to determine their future roles in clinical settings.

2.6 Positron Emission Mammogram

Clinical and small-animal positron emission tomography (PET) have expanded significantly during the past decade. PET is a molecular imaging technique that obtains images of molecular events in living subjects by localizing the radiotracers. The principle of PET imaging is based on the abnormalities of cellular function in tumors. Properly designed radiotracers can be used to probe tumor biology. The most commonly used PET radiotracer is ^{18}F-fluorodeoxyglucose (FDG) as tumor cells have a higher rate of glucose use. When injected into the bloodstream, FDG is taken up by the cancer cells at a higher rate than normal cells; thus, image contrast is obtained between tumor region with FDG and the neighboring tissues.

Previous studies using FDG PET to detect breast cancer showed that the sensitivity of FDG PET in detecting breast cancer was limited [68, 111]. PET may be helpful in breast cancer staging; however, previous study results are mixed and inconclusive [30, 68, 111]. Some studies have suggested that FDG

PET can be helpful in detecting distant metastases from breast cancer and in monitoring response to therapy [65, 68].

A dedicated breast PET scanner, also called positron emission mammography (PEM) has been developed in recent years. Gentle compression, in contrast to the firm compression in x-ray mammography, is applied to immobilize the breast during the PEM scan. PEM generates a 3D volume with a high-resolution ($2mm$) in-plane image matrix. Currently, the PEM scanners do not have accompanying CT systems for attenuation correction. Therefore current PEM system can only provide qualitative imaging instead of quantitative imaging that is available with conventional PET/CT scanners. Dedicated breast PET/CT scanners are still in the prototype stage.

2.7 Summary of Current Breast Imaging Modalities

Currently, x-ray mammogram is the primary breast imaging modality, and ultrasound and breast MRI are two most often used adjunctive modalities. The roles of other imaging modalities in breast cancer detection and diagnosis are rapidly evolving. Newer ultrasound methods such as 3D ultrasound, and nuclear medicine methods such as positron emission tomography (PET), are all exciting areas of development in breast imaging. Each of these alternatives are being sought in order to overcome at least one of the two major inherent limitations of x-ray mammography: the information loss of visualizing a 3D structure in 2D projections and the lack of functional insight regarding the

biological processes of the breast tissue imaged. In Table 2.2, we provide a comparison of several breast imaging techniques.

In the near future, mammography will still be the first line screening imaging choice. However, the inherent limitations of mammography as a radiographic planar image will still exist even in digital mammography, therefore it is necessary for other breast imaging techniques to accomplish more than current mammograph can provide. Note that breast tomosynthesis technique is still in early development stage and breast CT is still in its infancy, the efficiency and future role of these two techniques are still unclear yet. Breast MRI will be used in screening women at high risk for developing breast cancer, and will be routinely used for staging. In addition, breast MRI is used for monitoring treatment response and for surveillance for disease recurrence. In the next 10 years, we predict that X-ray mammography will still be the primary breast screening modality, and breast MRI will be increasingly used an adjunct to mammography. More clinical trial results are needed in order to determine the roles of DBT and breast CT in breast cancer diagnosis and staging.

\	Mammo	MRI	Ultrasound	Tomo	CT	PET
Sensitivity	70–90%	95-98%	2D	NA	NA	NA
Specificality	$< 50\%$	20-95%	2D	NA	NA	NA
Ionization	Yes	No	No	Yes	Yes	Yes
Dimension	2D	dynamic 3D	2D or 3D	3D	3D	3D
Principle	density	physiology	impendence	density	density	physiology
Compression	Yes	No	No	Little	No	Yes
Scan time	10 mins	30 mins	10 mins	20 mins	20 mins	20 mins
Scan Cost($\sim$)	\$200	\$1500	\$200	NA	NA	NA
Equipment Cost($\sim$)	\$200K	\$2000K	\$40K	\$ 300K	\$ 400K	\$500K

Table 2.2: Summary of breast imaging modalities

Chapter 3

Computational Image Analysis for Breast MRI

High cost and low specificity are two factors impeding wide use of breast MRI in current breast cancer care management. In addition to the scanner and material cost, high interpretation cost is another factor causing the high cost of breast MR exams. Since breast MRI generates 4D data, substantial effort is needed by radiologists in order to interpret the large amount of imaging data. Breast MRI is a relatively new technique, many radiologists do not have sufficient training or experience to interpret the breast MR images. Thus the specificity of breast MRI exam varies from site to site depending on the experience level or skill of radiologists. To facilitate the use of this advanced imaging modality for improving breast cancer care, there is a need to develop an image analysis system to help interpret breast MRI in a more accurate, efficient, and consistent way.

This chapter begins by reviewing the diagnostic guidelines in practice for interpreting breast MRI, then discusses the needs and the challenges that arise in developing a computational image analysis system for breast MRI applications. Then research goals and design methodologies of this dissertation are

outlined to prototype and evaluate the algorithms. The last section of this chapter gives a description of the breast MR imaging data used in our studies.

3.1 Diagnostic Criteria for Interpreting Breast MRI

There are no universal interpretation criteria for breast MRI; however, some guidelines have been proposed [60, 79, 95]. The American College of Radiology [79] provides a standardized lexicon to describe findings on breast MRI, and suggests some criteria for interpreting breast MRI examinations. Experienced radiologists have also summarized guidelines based on their own expertise [60].

The first step in interpreting a breast MRI is to determine the configuration of the lesion, namely if the lesion is (a) a mass, (b) a non-mass-like enhancement, or (c) a "focus" [60, 79].

A "mass" is defined as "a space occupying tumor that has three dimensions" [79]. A mass can be either a benign tumor (*e.g.,* fibroadenoma) or a malignant tumor (*e.g.* breast cancer). Generally, masses with irregular shapes are more likely to be malignant than those with round, oval, or lobulated shapes [60, 79]. The margin of a mass is also indicative of its pathological classification. The margin of the mass can be described as smooth, irregular, or spiculated. A spiculated margin is suggestive of a malignant tumor, while a smooth margin is suggestive of a benign lesion. However, the accuracy of the margin informa-

tion is subject to the spatial resolution of the MR image. Margin information extracted from a low-resolution image may not be as accurate as that from a high-resolution image [63]. In addition to shape and margin features, the internal enhancement pattern within the lesion provides important differential diagnostic information. The internal enhancement patterns can be described as: (a) dark/enhancing internal septations, (b) homogeneous, (c) heterogeneous, (d) rim enhancement, or (e) central enhancement. Generally, dark internal septations and homogeneous enhancement patterns are suggestive of benign lesions, while heterogeneous, rim enhancement, and enhancing internal septation are typical of malignant tumors [60]. Dynamic contrast-enhanced breast MRI not only provides internal enhancement patterns for further assessment of the lesion, but also provide powerful differential diagnostic criteria through kinetic time curve analysis. Previous studies have shown that the kinetic enhancement curves of malignant and benign lesions are different (see Fig. 3.2) [62]. In evaluating kinetic time curves, the early wash-in rate and the delayed-phase enhancement patterns are assessed in a small region of interest (ROI) containing the brightest enhancement on the first post-contrast image. It is believed that the early wash-in rate of malignant tumors is higher than that of benign lesions. The delayed phase enhancement pattern is another feature that can be used to distinguish benign from malignant lesions. A persistent pattern is suggestive of a benign lesion, while a washout pattern is suggestive of a malignant lesion. A plateau pattern may be found in both benign and malignant lesions [95].

A “non-mass-like enhancement” finding on breast MRI can be due to an intraductal tumor or a benign process, such as focal adenosis, hormonal stimulation of normal tissues, or inflammatory changes/mastits. Rarely, non-mass enhancement may be due to an invasive lobular cancer [60]. Unlike a “mass”, the shapes and borders of a “non-mass-like enhancement” cannot be described. The distribution pattern of the enhancement is an important differential indicator. The distribution pattern can be described as “*segmental*,” “*ductal*,” “*linear*,” “*focal area*,” “*regional enhancement*,” or “*diffuse enhancement*” (enhancement in multiple regions). Usually, *segmental* or *ductal* enhancement patterns are indicative of intraductal tumors. In addition to the enhancement distribution pattern, the internal enhancement pattern within a region of ”non-mass-like enhancement” is another differential diagnostic indicator. The internal enhancement pattern can be described as *homogeneous, heterogeneous, stippled, or clumped*. The *clumped* and *stippled* patterns are typically due to ductal carcinoma in situ. However, the internal enhancement patterns of a “non-mass-like enhancement” are less indicative than those of a “mass” finding [60]. Besides the distribution patterns and internal enhancement patterns, symmetry of a “non-mass-like enhancement” is another differential factor. Asymmetric non-mass-like enhancement findings are more likely to be malignant than symmetric ones. The kinetic time curves of ”non-mass-like enhancements” should be assessed in a manner similar to the way that masses are analyzed.

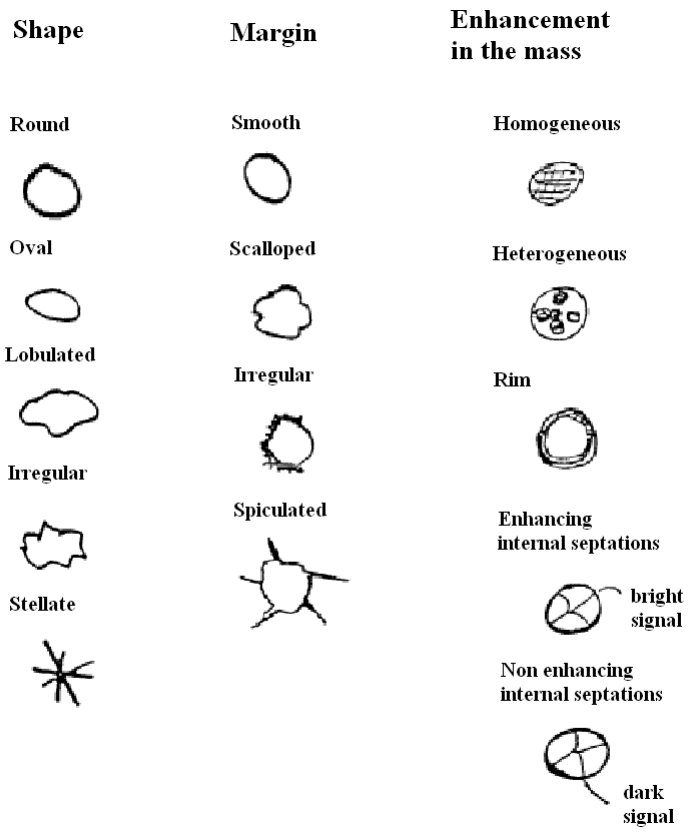


Figure 3.1: Mass Pattern. Reproduced from Fig. 3 of [95]

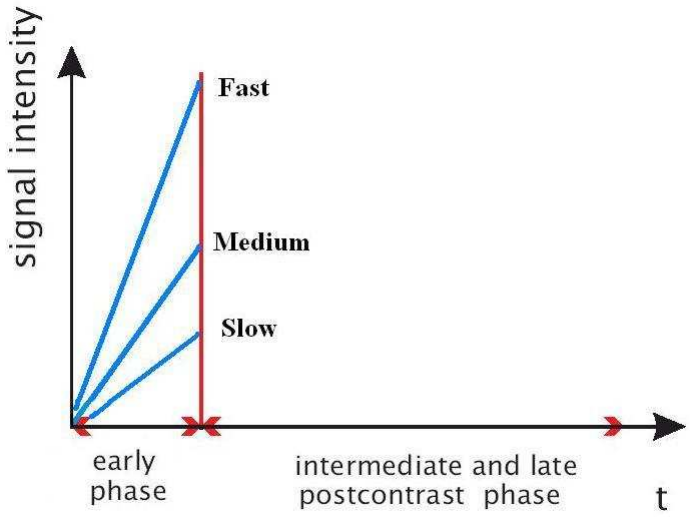
.

The third type of finding on breast MRI is a “focus”, which refers to a small enhancing area less than 5mm [79]. Due to its tiny size, the shape and margins of a "focus" can not assessed, and its kinetic information may be compromised due to partial volume averaging with its surrounding tissue [79]. Therefore, it is usually not possible to further classify "focus" findings on breast MRI [60, 79], and they must be further evaluated based on other data.

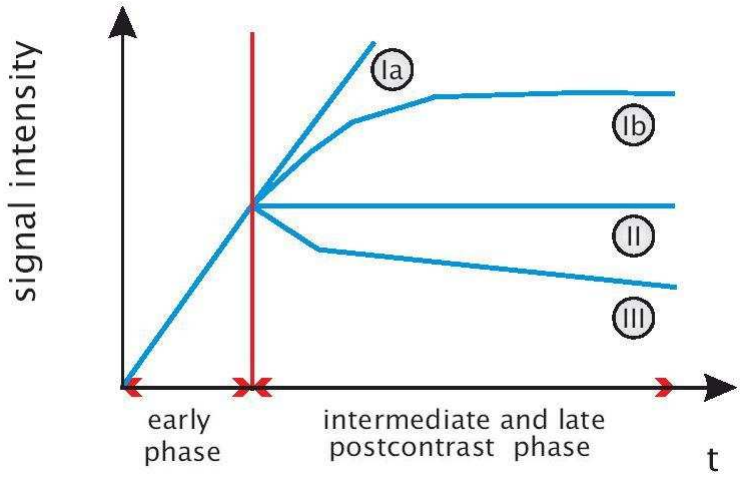
This dissertation project will employ mass lesions only since most cancers present as masses on breast MRI exams [109].

3.2 Computer-aided System for Breast MRI

The accuracy of image-based diagnosis is determined by both image acquisition and image interpretation. Consequently, the use of computer-aided detection (CADe) and computer-aided diagnosis(CADx) systems have been rapidly developed at several academic and industry sites for the purpose of aiding radiologists in interpreting medical images. CADe systems are intended to aid radiologists in the process of detecting and localizing abnormalities on medical images, and as such they help overcome perceptual errors such as oversight. CADx systems are intended to aid radiologists in the process of predicting the pathological nature of a lesion and to assist in selecting an appropriate next action, typically biopsy or follow-up imaging. Thus, CADx systems can help avoid missed cancers and obviate benign biopsies. It is important to rec-



(a) Early phase kinetic curve pattern



(b)Intermediate and late phase kinetic curve pattern

Figure 3.2: Ideal kinetic enhancement curve patterns of benign and malignant breast lesions on DCE MRI. The horizontal axis represents time and the vertical axis represents signal intensity. The curves labeled types Ia and type Ib are typical of benign lesions while those labeled type II and type III are typical of malignant lesions. (b) reproduced from [62]

ognize that CADe/CADx systems are designed to assist radiologists, not to perform autonomous diagnosis, although early experiments focused on the independent performance of CADe/CADx systems [113]. In breast imaging, the majority of work to date in CADe/CADx has focused on x-ray mammography [93] computer-aided detection techniques that use image analysis algorithms to detect suspicious areas, initial studies show CADe technology may reduce the number of missed cancers on mammography and can significantly enhance the performance of general radiologists [43, 78]. CADx systems for mammography are not yet in routine clinical use, but these systems have been the subject of numerous research studies.

The core of CADe/CADx systems is the computational image analysis algorithm(s), and the entire CAD system is the implementation of image analysis algorithms in a specific computer language. Two common medical image analysis modules are image segmentation and image registration. Image segmentation is to automatically identify relevant structures in the images. Image registration is to establish correspondence between the images acquired using different imaging modalities (inter-modal registration) or at different times (intra-modal registration). In addition, image classification and machine learning techniques can improve the specificity of abnormal findings by predicting that a detected lesion is benign or malignant. Other information, such treatment options, can be predicted based on the quantitative imaging features as well.

The high cost and varying specificity of breast MRI are two factors restricting its use in routine clinical practice. CADe/CADx systems for DCE breast MRI are needed to enable more efficient and accurate interpretation of breast MRI exams. In fact, improving efficiency is the primary goal of current commercial systems for computer-aided DCE breast MRI interpretation. To the best of our knowledge, there are currently four commercial software systems for aiding with DCE breast MRI interpretation that are approved by the FDA: CADstream (Confirma, Kirkland, WA), fTP (CAD Sciences, New York, NY), DynaCAD (Invivo, Orlando, FL), and OnCAD (Penn Diagnostics, Rockville, MD). While these systems represent an important step in breast MRI CADe/CADx development, one should recognize that they are currently more limited than their more established counterparts in x-ray mammography. In particular, all systems currently on the market are visualization software tools that aim at improving efficiency only. At the present time, no commercial system attempts to improve the specificity of breast MRI exams, and there are no standardized or automatic approaches for analyzing or interpreting breast MR images. Thus, there is a great need for a CADx system for DCE breast MRI.

3.3 Research Design and Methods

In order to develop an effective CADx system for breast MRI, research is needed on the necessary algorithmic components (including lesion segmentation, image registration, and lesion classification) and the development of common metrics and methods for evaluation. The goals of the this dissertation are to develop image analysis methods that ultimately would improve the performance of each of the major algorithm components that would be needed in a breast MRI CADx system. Diagnostic criteria used in current clinical practice suggest that both morphological and kinetic features of lesions be evaluated in making diagnosis and treatment decisions [60, 79, 95]. Thus, the lesion needs to be accurately segmented, not simply localized, so that the morphological features can be measured. Parallel to the lesion segmentation step, the kinetic features of the lesion need to be extracted accurately as well so as to provide additional differential diagnostic information. In order to extract accurate kinetic information, an image registration step is necessary to spatially align the voxels across sequentially collected breast MRI volumes to ensure accurate time curve signal representation at each spatial location of the lesion. With morphological and kinetic features quantitatively measured, a classification model can be constructed to determine whether a lesion is more likely to be benign or malignant.

A flow chart is described in 3.3 to illustrate the image analysis pipeline.

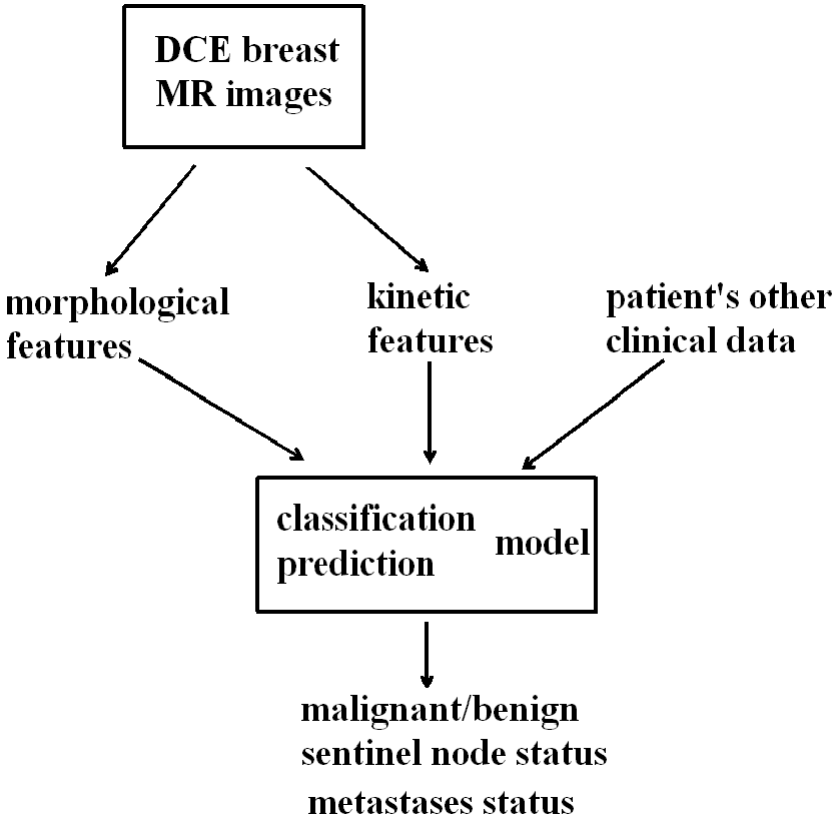


Figure 3.3: System flow chart

.

3.4 Data Set

The image database used in this study was collected retrospectively under a research protocol approved by institutional review boards of both The University of Texas at Austin and The University of Texas M. D. Anderson Cancer Center. Women between the ages of 21 and 75 years who underwent breast MRI examinations at The University of Texas M. D. Anderson Cancer Center between Arpil 01, 2006 and April 15, 2007 were eligible for this study. Images were obtained with fast spoiled gradient recalled sequence (TR/TE= milliseconds, echo time varies from 1.6 to 2 milliseconds, flip angle = 15°). Patients were scanned while in the prone position on a 1.5-Tesla MRI scanner (General Electric Medical Systems, Milwaukee,WI,USA). After acquisition of the precontrast volume image, a gadolinium-based contrast agent was injected intravenously by a power injector, and followed by a saline injection. Four postcontrast series were obtained, with an interval about 120 seconds. Each image matrix had 256 ×256 pixels and the slice thickness was 1.5 mm (3 mm thickness acquired, 1.5 mm spacing after slice interpolation).

A query search resulted in a total 363 patients with a total of 390 breast MRI exams. After reviewing the report from each MRI examination, we excluded patients from our study for any of the following reasons:

(1) There was no lesion found on the breast MRI examination;

(2) The patient received chemotherapy or radiation therapy prior to

the examination;

(3) The lesion displayed on the breast MRI was too small (smaller than 5 mm);

(4) The patient was diagnosed with inflammatory breast cancer;

(5) The patient stipulated that her medical records be reviewed only by those directly involved in the patient's care; (such cases were dropped immediately without reviewing the patients' imaging reports).

(6) The patient was unable to maintain the prone position during the examination, or the examination was interrupted due to the patient's discomfort, or the entire exam was incomplete.

We studied a total of 79 mass findings on breast MRI. Of these 79 masses, 50 were malignant, and 29 were benign.

We determined whether a lesion was malignant based on the following criteria: (1) MRI/ultrasound guided needle core biopsy proved the lesion to be malignant; or (2) Fine needle aspiration cytology proved the lesion be malignant;

We determined whether a lesion was benign based on the following criteria: 1) pathology results proved the lesion to be benign; or (2) follow-up mammograms, breast MRI, or ultrasound after at least twelve months showed the targeted lesion as stable and benign.

Chapter 4

Lesion Segmentation

Diagnostic criteria used in current clinical practice suggest that both morphological and kinetic features of lesions be evaluated in making diagnosis and treatment decisions [60, 79, 95]. The segmentation step is crucial because errors at this stage impact subsequent morphological feature extraction and classification. Thus, the lesion needs to be accurately segmented from the surrounding tissues, not simply localized, so that the morphological features can be accurately measured and evaluated.

This chapter concerns lesion segmentation on breast MRI. We begin by reviewing previous lesion segmentation work in breast MRI. In our study, we utilize a probabilistic method to obtain an optimal segmentation map based on several algorithmic segmentation outputs. In evaluating the segmentation performance, we further assess the impact of the resulting segmented lesions on subsequent classification performance.

4.1 Previous Segmentation Methods of Breast MRI

Image segmentation is a well-explored field. The purpose of image segmentation is to partition the image into disjoint sub-regions, and each resulting sub-region represents a different object. Several previous publications are directly related to lesion segmentation on breast MRI. Arbach *et al.* [4] segmented the subtraction image, which is generated by subtracting the pre-contrast image from the first post-contrast image. The segmentation in Hayton's work was based on thresholding the enhancement rate of the first two images in the series [50]. Gihuijs *et al.* [45] developed a seeds-based volume-growing algorithm to segment the lesion from the ROI (region of interest) using thresholds derived from the image histogram. Petroudi *et al.* [86] proposed a Gaussian mixture model for dynamic breast MRI data, in which the relationships between voxels are modeled by a Markov random field (MRF) and the mixture parameters are iteratively updated based on K-means initialization. However, K-means initialization is not stable and not recommended in general. Chen *et al.* [21] proposed an automatic lesion segmentation method based on a fuzzy c-means technique to partition all voxels within the ROI to either lesion or non-lesion classes, and morphological processing was applied to generate a smooth segmentation map. Wu *et al.* [114] proposed an interactive lesion segmentation method by generating a maximum *a posteriori* (MAP) estimate of the class membership (lesion vs. non-lesion) for each voxel within the ROI. The prior distribution of the class membership in [114] was modeled as a multi-level logistic model, a MRF model in which the class membership of each voxel is

assumed to depend upon its nearest neighbors only.

Among these methods, the segmentation is performed on raw breast MR sequences directly in [4, 21, 45]; thus, the segmentation results are compromised by the motion displacement across the image series. Though the lesions are segmented on the motion-corrected images in [50, 114], the performance of image registration is difficult to validate and it is not expected to compensate completely for all motion distortions. Since the segmentation algorithms in [4, 21, 45, 50, 86, 114] use two or more image sequences of the raw or the registered MRI sequences, the segmentation results are compromised, more or less, by motion distortion.

4.2 The Segmentation Scheme in Our Study

Lesions, along with other tissues, such as ducts, lymph nodes, and blood vessels, enhance on post-contrast images [71]. Thus, enhancement alone may lead to a false positive detection; therefore a more sophisticated algorithm would be required for lesion detection. Rather than attempting to automate the detection step, we assume that the ROI has already been detected by the human reader. In our study, we only investigate segmenting the lesion from its surrounding tissues within the manually pre-defined ROIs.

Since DCE breast MRI generates 4D imaging data, one may choose to segment

lesions using 4D. However, due to motion, the same location at different times may not be associated with the same physical object. In order to eliminate the impact of motion distortion on segmentation performance, we perform segmenting the lesion from the surrounding tissues on the first post-contrast image, which also provides strong contrast between the lesion and the non-lesion tissues. This practice is consistent with the American College of Radiology's guidelines [79] which suggest that lesion measurement be performed on the first post-contrast image.

4.2.1 STAPLE Algorithm to Optimize the Segmentation Map

A single segmentation algorithm may not prove to be consistently superior to all other methods given the complexity of breast MR images. If a set of automatic segmentation generators provide complementary information, it may be beneficial to combine their assessments. Since a large number of image segmentation algorithms have already been developed, this study focuses on optimizing the segmentation results by computing a probabilistic segmentation based on a collection of algorithmic segmentation maps, such as those generated by threshold based segmentation, the fuzzy C-means segmentation algorithm, and the mean shift segmentation algorithm.

Recently the STAPLE (simultaneous truth and performance level estimation) algorithm was introduced to combine segmentations from multiple human

raters or automated algorithms in order to compute a probabilistic estimate of the true segmentation map [106]. In the STAPLE formulation, the expert segmentation decision at each voxel is known information, and the hidden true segmentation is a binary variable but unknown for each voxel. The performance level, or quality, achieved by each segmentation (either human or algorithm) is represented by sensitivity and specificity parameters. The STAPLE algorithm estimates the unknown true segmentation map by using the expectation- maximization (EM) algorithm. Detailed formulation of the STAPLE algorithm will be discussed in the following paragraph.

The notations here follow the ones in the paper by Warfield *et al.* [106]. Consider an image of N voxels, and the task of segmenting a structure in that image by indicating the presence or the absence of the structure (lesion in our study) at each voxel. Let $p = (p_1, p_2, \ldots, p_R)^T$ be the sensitivity parameter, with each element characterizing one of R segmentations; and likewise $q = (q_1, q_2, \ldots, q_R)^T$ be the specificity parameter. Let matrix $D_{N\times R}$ denotes the binary decisions made by each segmentation at each voxel of the image. The hidden binary true segmentation vector, $T_{N\times 1}$, denotes each voxel as 1 (lesion) or 0 (non-lesion).

The STAPLE method formulates a probabilistic estimate of the true segmentation as an optimal combination of the segmentation results D. Then the goal is to find the optimal performance level parameters (p^*, q^*) that maximize the

following complete data log likelihood function:

$$(p^k, q^k) = \arg\max_{p,q} E[ln\frac{f(D,T,p,q)}{f(p,q)}|D, p^{k-1}, q^{k-1}] \tag{4.1}$$

The performance parameters, and a probabilistic estimate of the true segmentation, can be optimized using the EM algorithm iteratively. The first step of each iteration is to estimate the conditional probability of the true segmentation given the expert decisions and the previous performance parameter estimates (see Equation (4.2)), and the second step updates the estimation of the performance parameters (see Equation (4.3)).

$$a_i^k = f(T_i = 1)\prod f(D_{ij}|T_i = 1, p_i^{(k)}, q_i^{(k)}) = f(T_i = 1)\prod p_j^{(k)}\prod(1 - p_j^{(k)}) \tag{4.2a}$$

$$b_i^k = f(T_i = 0)\prod f(D_{ij}|T_i = 0, p_i^{(k)}, q_i^{(k)}) = f(T_i = 0)\prod q_j^{(k)}\prod(1 - q_j^{(k)}) \tag{4.2b}$$

$$W_i^{k-1} = f(T_i = 1)|D_i, p^{(k-1)}, q^{(k-1)}) = \frac{a_i^{(k-1)}}{a_i^{(k-1)} + b_i^{(k-1)}} \tag{4.2c}$$

$$p_j^k = \frac{\sum_{i:D_{ij}=1} W_i^{k-1}}{\sum_i W_i^{k-1}} \tag{4.3a}$$

$$q_j^k = \frac{\sum_{i:D_{ij}=0} 1 - W_i^{k-1}}{\sum_i 1 - W_i^{k-1}} \tag{4.3b}$$

4.2.2 The Automatic Segmentation Algorithms

As described above, the STAPLE algorithm is based on several manual or algorithmic segmentation results. The segmentation maps on which the STAPLE algorithm utilizes are obtained by the following segmentation methods:

1. **Threshold-based segmentation**: Global thresholding is a simple approach of image segmentation if there are multiple clearly separated peaks in the image histogram. This method labels each pixel of the image based on the intensity value. Since the lesion is enhanced after the injection of the contrast agent, threshold based methods could be used for our application. We used an automatic method [81] to determine the threshold that maximizes the separability of the resulting sub-regions. This method is nonparametric and unsupervised in the sense that no *a priori* knowledge about the image is needed.

In our experiments, the voxels with intensity value above the threshold are considered as the lesion. However, segmentation on the basis of intensity alone lacks spatial coherency. There maybe some small holes inside the segmented lesion caused by necrosis or image inhomogeneity. A morphological operation is performed to fill in these small holes.

2. **Fuzzy C-means Segmentation**

The Fuzzy C-means (FCM) algorithm was formulated by modifying the objective function of the standard FCM algorithm to compensate for intensity inhomogeneities and to allow the labeling of a pixel (voxel) to be influenced by the labels in its immediate neighborhood. The neighborhood acts as a regularizer and biases the resulting segmentation map toward piecewise-homogeneous labeling; such a regularization is useful in segmenting scans corrupted by salt and pepper noise. In our experiments, the initial segmentation map was obtained by the automatic threshold algorithm discussed above.

The basic idea of the FCM algorithm can be summarized below. The standard FCM partitions the data $\{x_k, k = 1...n\}$ into c groups by minimizing the following objective function:

$$J = \sum_{i=1}^{c} \sum_{k=1}^{N} u_{ik}^{p} \|x_k - v_i\|^2 \tag{4.4}$$

where where v_i are the cluster centers and the partition matrix u_{ik} denotes a likelihood of data x_i belonging to the cluster i and the elements are subjected to the following constraint $\sum_{i=1}^{c} u_{ik} = 1, 0 \leq u_{ik} \leq 1$. In [3], Ahmed *et al* proposed a modification to (4.4) by introducing a term that allows the labeling of a pixel (voxel) to be influenced by the labels in its immediate neighborhood [21]. The modified objective function is formulated as follows:

$$J = \sum_{i=1}^{c} \sum_{k=1}^{N} u_{ik}^{p} \|x_k - v_i\|^2 + \frac{\alpha}{N} \sum_{i=1}^{c} \sum_{k=1}^{N} u_{ik}^{p} (\sum_{x_r \in \mathcal{N}_k} \|x_k - v_i\|^2) \tag{4.5}$$

where $\mathcal{N}_k$ denotes data α, and N_R is the cardinality of the window $\mathcal{N}_k$.

3. **Mean-shift Segmentation**

Ideally, spatial information should be incorporated into a segmentation scheme. After pre-processing an input image to be segmented, a pixel/voxel in the image can be mapped into a point in the feature space (joint spatial-intensity space). Significant features will be grouped together and form a denser region, namely a cluster, in the feature space. In this manner, image segmentation can be posed as a clustering task in the feature space that is derived from the input image [28].

Mean shift is a data-clustering method that searches the modes of the underlying probability density and clusters the data without estimating density explicitly [23, 26, 27, 40]. This method is based on density estimation theory and its efficient gradient estimation [35, 38, 84, 97]. Even though the mean shift method was first studied in 1975 [38], only recent work [23, 26, 27] has exploited the properties of mean shift for clustering and image segmentation. The mean shift method has been demonstrated to be successful for several segmentation tasks [26–28]. Here we discuss briefly the basic principles of the mean shift method, and we refer the interested readers to [23, 26, 27] for more details.

Assume that for each data point $\mathbf{x}_i \in R_d, i = 1, \ldots, n$, the sample point

estimator at location $\mathbf{x}$ can be written as

$$\widehat{f}(\mathbf{x}) = \frac{1}{nh^d} \sum_{i=1}^{n} K(\frac{\mathbf{x} - \mathbf{x}_i}{h}) \tag{4.6}$$

where the kernel $K(\mathbf{x})$ is the multivariate kernel and h is the *window bandwidth*, also called the *smoothing parameter*. A large value of h provides the global structure of the data, while small window radii h reveal the local structure. Previous experiments have demonstrated that the segmentation is not very sensitive to the choice of bandwidth h [28].

In our experiments, each breast MRI image is represented as a 3-dimensional lattice of scalars (voxels). The space of the lattice is referred as the spatial domain while the image intensity information is referred to as the range domain. In implementing the mean shift algorithm for lesion delineation on breast MRI, we want to group into the same cluster only those voxels that are both close in the spatial domain and similar in image intensity. Thus, the spatial and the range domain information can be concatenated to generate a joint spatial-range (or spatial-intensity) domain [26], namely the scalar value $I(x, y, z)$ in the range domain for a given voxel (x, y, z) is transformed to a 4D vector $[x, y, z, I(x, y, z)]$ in the joint spatial-range domain. The mean shift method is applied to the joint spatial-range vectors of all voxels within the ROI to find their modes. The voxels sharing the same mode are partitioned into the same sub-region; thus, the ROI is partitioned into different sub-regions.

4.3 Evaluation of Segmentation Algorithm

Three factors, precision (reproducibility), accuracy (agreement with truth, validity), and efficiency (time taken), are usually considered in evaluating segmentation methods [103]. In this dissertation, we emphasize on accuracy in evaluating segmentation performance.

Note that the ultimate goal of segmentation is to to extract morphological features that will be used as input features for the classifier or prediction model; therefore more relevant in our study is to assess the impact of the extracted morphological features from segmented lesions in subsequent classification stage.

4.3.1 Evaluation Metrics for Segmentation Accuracy

The quality of the lesion segmentation is traditionally evaluated using either subjective or objective methods. The subjective evaluation by radiologists, such as a five-level rating, is considered as the gold standard in clinical practice. Without the known ground truth data obtained by non-imaging methods such as histology, objective assessment of the accuracy of a segmentation algorithm requires a surrogate for ground truth and an appropriate metric for comparing the segmentation to the gold standard. Currently, the only accepted surrogate for ground truth in the medical imaging community is manual segmentation by human experts. Compared to the subjective evaluation, the objective metrics are easy to compute, but may not provide an overall assessment of the

segmentation quality because each metric only captures a certain aspect of the difference between the segmentation and the ground truth. Although the study of methodology for assessing segmentation accuracy remains an active area of research, there are several commonly used metrics, such as the volumetric overlapping ratio, the Hausdorff distance, and the mean absolute surface distance [41], that have proven useful in practice.

The overlapping ratio is a simple yet meaningful measurement; thus, we will employ the overlapping ratio in our experiments to assess segmentation accuracy. There are several ways to define the overlapping ratio, such as the Dice coefficient and the Jaccard coefficient. The Dice coefficient is defined as the ratio of twice the size of the intersection of two sets divided by the sum of two individual set sizes, as formulated in 4.7. A simple spatial overlap index is the Dice similarity coefficient (DSC), first proposed by Dice (13). The Dice similarity coefficient is a spatial overlap index and a reproducibility validation metric. The Jaccard coefficient is defined as the size of the intersection of two sets divided by the size of the union of two sets (4.8).

$$Dice = 2\frac{|A \cap B|}{|A| + |B|} \tag{4.7}$$

$$Jaccard = \frac{|A \cap B|}{|A \cup B|} \tag{4.8}$$

In (4.7) and (4.8), $|.|$ denotes the number of pixels/voxels in a given set; $A \cap B$

denotes the intersection of set A and set B; while $A \cup B$ denotes the union of set A and set B. There is an explicit relationship between the Dice coefficient and the Jaccard coefficient. Therefore there is no difference between Dice coefficient and Jaccard coefficient in characterizing the accuracy.

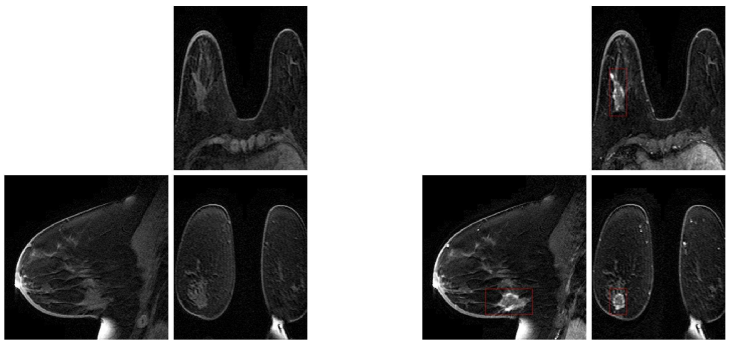
4.3.2 Evaluation of Segmentation Impact on Classification Performance

The metrics are indicators of the segmentation performance. Our hypothesis is that more accurate lesion segmentation will result in more accurate lesion classification. More relevant about the segmentation performance in our study is the subsequent classification performance resulting from the features derived from the segmented lesion. A set of morphological features from the segmented lesion can be extracted and their classification performance then will be evaluated. This evaluation method provides an indirect but more meaningful way in assessing the segmentation methods.

In our study, we consider the manually segmented lesion as the "ground truth" lesions, and the morphological features can be extracted from the segmented lesions. Then a subset of features can be determined by the feature selection method. The same subset of features will used in evaluating the classification performance of other segmentation algorithms.

4.4 Experimental Results

We carried out lesion segmentation on 40 lesions (20 malignant and 20 benign). Some experimental results are shown in Fig. 4.1 to Fig. 4.4. The images shown in Fig. 4.1 are the pre-contrast images and the 1st post-contrast images from a patient diagnosed with a biopsy-proven invasive ductal carcinoma, and each 3D image is $208 \times 256 \times 256$ in size. The ROI size of this case is $25 \times 37 \times 67$, as denoted by a red box in Fig. 4.1(b). The images shown in Fig. 4.2 are enlarged sagittal views of the ROI and segmentation results by different algorithms. The images shown in Fig. 4.3 are the pre-contrast images and the 1st post-contrast images from a patient diagnosed with a biopsy-proven benign fibroadenoma; in this case each 3D image is $168 \times 256 \times 256$ in size and the ROI size is $11 \times 18 \times 18$.



(a) Pre-contrast Image (b)First post-contrast Image

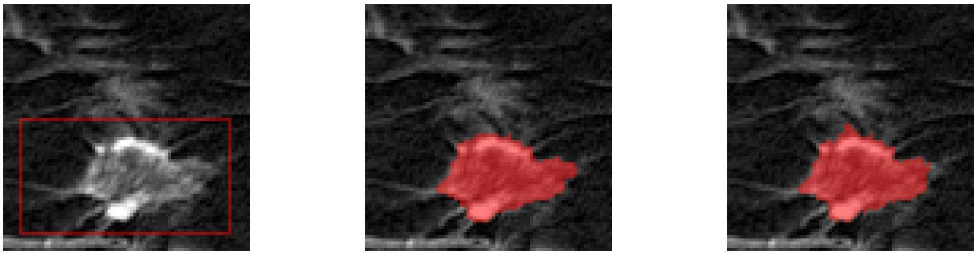
Figure 4.1: Axial, coronal, and sagittal views of raw 3D images from a patient of 46 years old diagnosed with invasive ductal carcinoma.

4.4.1 Results on Segmentation Accuracy

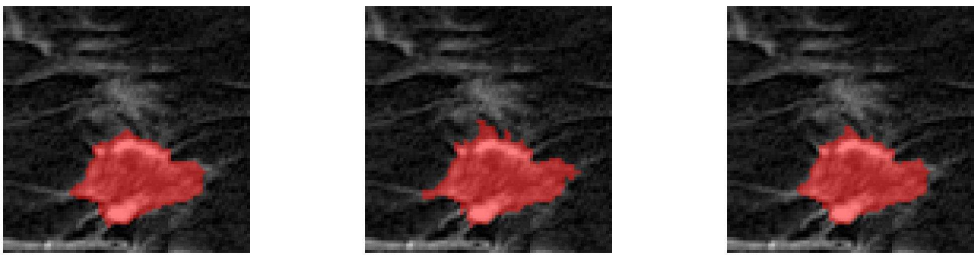
To evaluate the segmentation algorithm performance, we manually segmented out 40 lesions (20 malignant cases and 20 benign cases) as the "ground truth" and compared the algorithmic outputs with the manual results. Three metrics; sensitivity, specificity, and overlapping coefficient; were assessed, and the numerical results are summarized in Table 4.1. Overall, the STAPLE algorithm outperforms other algorithms in terms of all three metrics except that the threshold-based algorithm performs best in term of sensitivity and in terms of the Dice coefficient for malignant lesions. The performance of the Fuzzy C-means algorithm is comparable to that of the mean-shift algorithm.

4.4.2 Segmentation Impact on Classification Performance

In our study, we evaluated lesion segmentation by its impact on the subsequent classification performance. Similarly, the manually segmented lesions are used

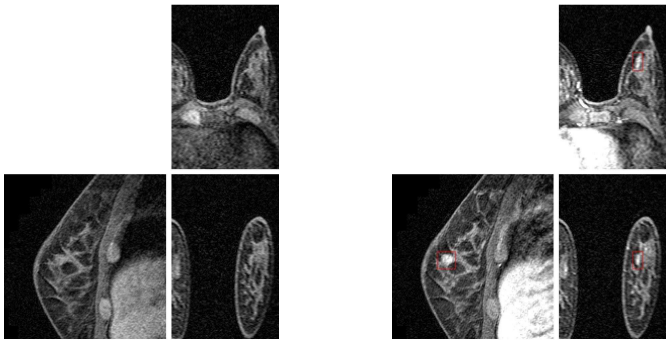


(a) 1st Post image (b) Manual map (c) Threshold-based seg



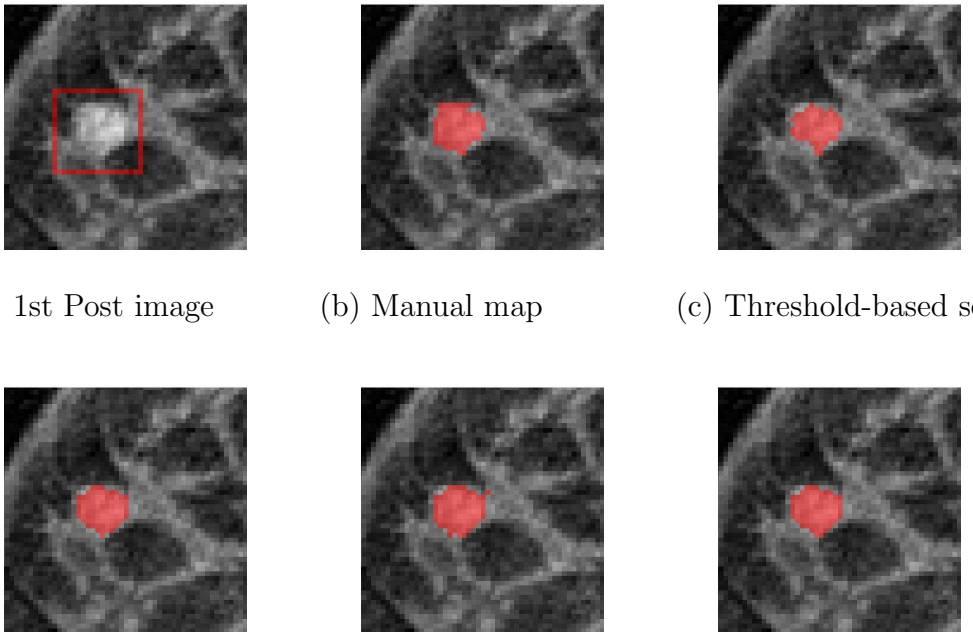
(d) Fuzzy C-means (e) Mean shift (f) STAPLE

Figure 4.2: Enlarged sagittal view (slice #43) of Fig.4.1(b) and the segmented lesions.



(a) Pre-contrast Image (b)First post-contrast image

Figure 4.3: Axial, coronal, and sagittal views of raw 3D images from a patient of 42 years old diagnosed with a benign fibroadenoma.



(a) 1st Post image (b) Manual map (c) Threshold-based seg

(d) Fuzzy C-means (e) Mean shift (f) STAPLE

Figure 4.4: Enlarged sagittal view (slice #124) of Fig.4.3(b) and the segmented lesions.

	Sensitivity (B/M/All)	Specificity (B/M/All)	Dice Coefficient (B/M/All)
Threshold	0.976/0.970/0.974	0.981/0.994/0.986	0.716/0.905/0.783
Fuzzy C-means	0.868/0.862/0.866	0.991/0.990/0.991	0.757/0.798/0.772
Mean shift	0.831/0.944/0.871	0.992/0.990/0.991	0.739/0.831/0.772
STAPLE	0.966/0.967/0.967	0.992/0.993/0.992	0.848/0.899/0.866

Table 4.1: Segmentation accuracy for different algorithms compared against manual "groundtruth"; B–benign lesions; M–malignant lesions; All–benign and malignant lesions together.

as "ground truth" in the experiments. The lesion features were automatically extracted to quantify the characteristics of the lesion. A subset of features were selected by the feature selection methods described in Section 6.2. The same selected features of lesions, both manually segmented and maps segmented out by algorithms, were evaluated using receiver operating characteristic (ROC) analysis. The area under the ROC curve is used as the performance metric.

The classification results are summarized in Table 6.2. Based on the algorithmic segmentation outputs of 40 lesions (20 benign and 20 malignant each), the STAPLE and the Mean Shift algorithms yield better classification results than the Fuzzy C-means and the threshold-based methods.

Chapter 5

Breast MR Image Registration

It usually takes around 20-40 minutes to scan a series of breast MR images. During this relatively long acquisition process, respiratory and cardiac motions, as well as some degree of voluntary patient movement, are unavoidable. As a result of such motion, the same coordinates in images at different times in the series may correspond to different physical locations in the subject. Thus, interpreting raw images can lead to errors in evaluating enhancement and morphology of an abnormality. For this reason, one must compensate for the motion across image volumes acquired sequentially in order to achieve anatomical and functional correspondence. This process is referred to as registration.

In this chapter, we begin by introducing the general methodology of image registration, then followed by reviewing previous image registration methods for breast MRI application. In our study, we conducted experiments by comparatively assessing the similarity metric in registering DCE breast MR images. The performance of cross correlation(CC) coefficient, and mutual information (MI) are studied in both rigid and non-rigid registration schemes. Numerical

results and statistical properties are reported. The resulting image quality after registration is discussed both qualitatively and quantitatively.

5.1 Overview of Image Registration

Registration is an essential tool in medical imaging. Because of the great variety of registration techniques, here we only review image registration techniques relevant to our application, breast MRI registration. We refer interested readers to recent publications [48, 53, 87] for broader and more detailed surveys on this topic. Suppose we have two images, a reference image, R, and a template image, T, both of which are mappings from d-dimensional space to $1D$, i.e., $R(\mathbf{x}), T(\mathbf{x}) : \Omega := \Re^d \rightarrow \Re$ ($d = 3$ in our application). The goal of image registration is to establish correspondence between voxels in these two images R and T. In general all image registration methods can be categorized as either *feature-based* or *intensity-based*. The *feature-based* methods are based on *fiducial markers* manually placed in the two images or salient features extracted from the images, such as edges. Once the correspondence of feature points is established, the two images can be aligned. In breast imaging, landmarks may be attached to the skin or inserted inside the breast. Those attached to the skin are non-invasive; however, the movement of skin may cause landmarks to move several millimeters [48]. Landmarks placed inside the tissue can cause discomfort or damage to the tissue [48]. Thus, fiducial markers are not used in most current clinical breast MR image acquisitions. Another way is to identify landmarks on images either manually by an operator or automatically by a

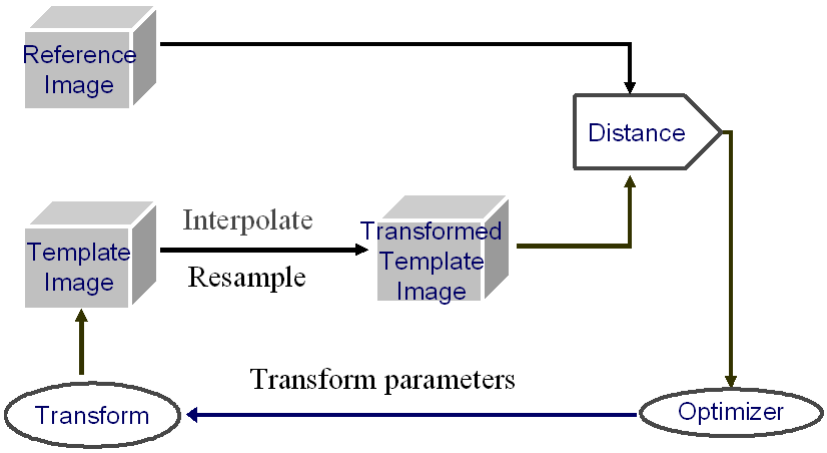


Figure 5.1: The iterative process of intensity-based registration.

computer. Manually identifying corresponding landmarks on breast MR image pairs is a challenging task even for experienced radiologists. Thus our study focuses on *intensity-based* registration methods.

Intensity-based methods involve an iterative process, in which an initial estimate of the transform is gradually refined (see Fig. 5.1):

(1) Generate a transformed template image based on the current transformation. Given a set of transformation parameters, each grid point in the transformed template image may correspond to a non-grid point in the original template image; thus, interpolation is necessary to evaluate the image intensity on those non-grid points;

(2) Measure the similarity or closeness of the transformed template image to

the reference image;
(3) If the closeness satisfies a pre-defined criterion, the iteration stops; otherwise, the optimizer searches for a new set of transformation parameters and steps 1-3 are repeated;
Each of these steps will be discussed in detail in the following paragraphs.

In general, transforms can be categorized as one of the following three types: rigid transforms, affine transforms, or non-linear transforms. The rigid transform only involves rotations and translations. In 3D, rigid transformation has 6 DoF (degree of freedom): three rotations (one about each axis) and three translations. The rigid transform does not allow any structure within the image to change size or shape. In comparison, the affine transformation model permits scalings and skews in addition to rotations and translations; therefore, it has 12 DoFs: 3 rotations, 3 translations, 3 scalings, and 3 skew parameters. Rigid transformation is a special case of affine transformation. Although affine transform allows both size and shape changes, the changes must be the same across the entire image. Mathematically, the affine transform can be described in a linear form. Transforms that are not affine are referred to as non-linear transforms.

A variety of non-linear transforms have been proposed based on different mathematical or physical models. The elastic model [5, 39] and the fluid model [24] are the most advanced registration methods available. The elasticity model

assumes linear elasticity of the soft tissue, allowing the relative positions of physical tissue points to be altered. This relative position change of physical points is referred as *deformation*. An elastic model deforms the image to reach the desired deformation by balancing between internal strain and external force. The internal strain is intrinsic to the elastic model while the external force is the similarity metric to be discussed later in this section. Due to the intrinsic internal force, the elastic model constrains the deformation to be globally smooth. Thus, the elastic model can not accommodate large, complex deformations. To overcome this limitation, the fluid model [24] has been proposed to control the deformation in which the internal force term disappears gradually so that the desired deformation can be reached. The original implementation by Christensen [24] is computationally intensive, but fast implementation techniques have been proposed by other researchers [13, 110]. However, fluid registration introduces noticeable blurring artifact to the transformed image.

After the template image T is transformed according to the chosen transform, it needs to be re-sampled to generate an intermediate image, then this discrete image is compared against the reference image R. The re-sampling process necessarily involves interpolation. The ideal interpolation kernel is the *sinc* function. However, the *sinc* interpolation kernel is impractical to implement due to its infinite support. Several other interpolation kernels have been investigated, such as various order B-spline interpolation kernels. Linear (order

1 B-spline kernel) interpolation causes errors though it is cheap to compute. Netsch *et al.* [74] demonstrated that cubic and quintic B-spline interpolation are favorable in terms of computational cost and consistency of registration results. Wu *et al.* [112] also demonstrated that B-spline kernels are superior to other types of interpolation kernels in terms of interpolation accuracy and computational cost.

The three most common distance metrics used to measure the similarity or closeness of the reference image to the transformed template image are: (1) L_2 norm, namely, the sum of squared difference (SSD), (2) correlation coefficient (CC), and (3) mutual information (MI). The mathematical definitions of these three metrics are as follows:

$$SSD = \frac{1}{N}\sum_{j} |\mathrm{I}_{\mathrm{A}_j} - \mathrm{I}_{\mathrm{B}_j}|^2 \tag{5.1}$$

$$CC = \frac{\sum_j (\mathrm{I}_{\mathrm{A}_j} - \overline{\mathrm{I}_{\mathrm{A}}})(\mathrm{I}_{\mathrm{B}_j} - \overline{\mathrm{I}_{\mathrm{B}}})}{[\sum_j (\mathrm{I}_{\mathrm{A}_j} - \overline{\mathrm{I}_{\mathrm{A}}})^2 \sum_j (\mathrm{I}_{\mathrm{B}_j} - \overline{\mathrm{I}_{\mathrm{B}}})^2]^{\frac{1}{2}}} \tag{5.2}$$

$$MI = H(\mathrm{I}_{\mathrm{A}}) + H(\mathrm{I}_{\mathrm{B}}) - H(\mathrm{I}_{\mathrm{A}}, \mathrm{I}_{\mathrm{B}}) \tag{5.3}$$

$$H(\mathrm{I}_{\mathrm{A}}) = -\sum_{a} P_{\mathrm{I}_{\mathrm{A}}}(a) log P_{\mathrm{I}_{\mathrm{A}}}(a) \qquad \forall \mathrm{I}_{\mathrm{A}}(\mathbf{x}) = a \tag{5.4}$$

$$H(\mathrm{I}_{\mathrm{B}}) = -\sum_{b} P_{\mathrm{I}_{\mathrm{B}}}(b) log P_{\mathrm{I}_{\mathrm{B}}}(b) \qquad \forall \mathrm{I}_{\mathrm{B}}(\mathbf{y}) = b \tag{5.5}$$

$$H(\mathrm{I}_{\mathrm{A}}, \mathrm{I}_{\mathrm{B}}) = -\sum_{a}\sum_{b} P_{\mathrm{I}_{\mathrm{A}}\mathrm{I}_{\mathrm{B}}}(a,b) log P_{\mathrm{I}_{\mathrm{A}}\mathrm{I}_{\mathrm{B}}}(a,b) \qquad \forall\ \mathrm{I}_{\mathrm{A}}(\mathbf{x}) = a\ \&\ \mathrm{I}_{\mathrm{B}}(\mathbf{y}) = b \tag{5.6}$$

I_A and I_B are two images to be compared, and N is the total number of pixels/voxels in the overlapping region of the two images. $\overline{I_A}$ and $\overline{I_B}$ are the mean image intensity values of images I_A and I_B, respectively. $H(I_A)$ is the entropy of image I_A, and $H(I_B)$ is the entropy of image I_B. Likewise, the term $H(I_A, I_B)$ is the joint entropy of two images I_A and I_B, which reflects how much the joint histogram of two images disperses. From 5.3, we can see that maximization of MI is equivalent to minimization of joint entropy $H(I_A, I_B)$. The more similar the two histograms P_{I_A} and P_{I_B} are, the less dispersion the joint histogram $P_{I_A I_B}$ shows, and the lower the value of the joint entropy, $H(I_A, I_B)$. More details about MI and its application in image registration are available in a recent review article [87].

Appropriate selection of a distance/similarity metric is critical to the performance of an image registration method. In terms of computational cost, SSD is the cheapest metric while correlation and MI are more computationally intensive. In term of applicability, SSD can be used when there is no or little intensity variation for the same tissue across an image pair; therefore, SSD is often used for intra-modality registration. If the image intensities for the same tissue in two images are significantly different or for inter-modality registration, SSD is a poor metric choice while CC or MI can be used instead. From the definitions of CC and MI, it is clear that CC is optimal when there is a linear relation between the intensities across an image pair, and mutual information is maximized when the histograms of the two images are very sim-

ilar to each other. Therefore, both CC and MI can be used for inter-modality registration.

5.2 Previous Work on Breast MRI Registration

To our knowledge, no previous breast MRI registration algorithms involve landmarks. Thus, all previous work on this subject falls into the category of *intensity-based image registration.* Although *intensity-based image registration* is a traditional, well-explored topic in image analysis and computer vision, existing techniques cannot be simplistically applied to breast MRI. For example, some registration algorithms assume that the image content does not vary much across the images to be registered, but this is not the case for DCE breast MRI. In fact, the key feature of DCE breast MRI is the enhancement of abnormal tissues after the administration of the contrast agent. Since breast tissues are soft tissues, the motion of different tissues may not be the same, so a non-linear transformation or a non-rigid motion model is preferred, as also discussed by Rueckert *et al.* [91]. Additionally, breast MRI is a 3D exam with high spatial resolution, so true 3D registration techniques would be preferred. Rueckert *et al.* [91] introduce a non-rigid 3D registration method for breast MRI based on free-form deformations. The local motion model is a free-form deformation model based on B-splines to analyze the motion of the breast, and the choice of cubic B-splines produces a smoothly varying displacement field. In this scheme, a normalized mutual information criterion is chosen to

maximize the similarity between the pre-contrast image and post-contrast image. It should be recognized that the image intensities of certain regions, *e.g.*, tumors and blood vessels, vary over time due to the wash-in and wash-out of the contrast agent. This intensity change inevitably leads to differences between the histograms of any volume pair in a dynamic series, even if the two volumes are perfectly aligned. Moreover, the dispersion of the joint histogram depends largely on the size of the enhancement region. If the enhancement region is small, the overall joint histogram has little dispersion. In contrast, if the enhancement region is substantial compared to the entire image, one should expect large dispersion in the joint histogram and thus the joint entropy will not be small.

To accommodate the fact that the image intensity of the tumor region in breast MRI changes with time, the standardized compartmental model (see Fig. 2.4) is widely used to derive physiological parameters of tissue underlying DCE MRI. This modelling process involves the following steps:
(1) Convert the image intensity to T1 values;
(2) Convert the dynamic T1 values to contrast agent concentration;
(3) Estimate the contrast agent concentration in plasma, this is the so called *arterial input function* (AIF) estimation;
(4) Use the standardized two-compartment model for parameter estimation based on the diffusion of the contrast agent between the plasma and the tumor cells.

In step (1), there may exist an analytical relationship between the $T1$ value and image intensity depending on the hardware and the MRI sequence choice. Step (2) is to measure the contrast agent concentration from the dynamic $T1$ values (MRI signals). Since the addition of the contrast agent only alters the MRI signals indirectly by changing the relaxitivity of the water surrounding the contrast agent, the exact effect is not fully understood. Previous studies assume that the T1 value changes linearly with the change in the concentration of the contrast agent. However, recent research indicates that this relationship is non-linear and dependent on the intrinsic tissue type. Regarding step (3), several AIF models have been proposed, but experiments indicate that these models do not fit all individual studies well. Another alternative is to calculate the AIF directly from the image intensity, which involves steps (1) and (2). Very often, the assumptions made may not be valid for every tumor type, as discussed by Collins and Padhani [25]. Furthermore, it is difficult to validate the accuracy of the model-based parameters as there is no existing clinical standard.

Hayton *et al.* [51] use a modified Horn and Schunck optical flow model to minimize the fitting error of a two-compartment model. The consistency of the pharmacokinetic model fitting is used as the similarity criterion in registration. Buonaccrosi *et al.* [14] proposed to register the raw image and a synthetic image of the same time point, with the synthetic images generated by a two-

compartment model. As discussed above, model failures can occur and would decrease the overall registration accuracy.

5.3 Comparative Assesment of Similarity Metric

5.3.1 Study Goal and Experiment Designs

As discussed in Section 5.1, it is a difficult choice in deciding the best registration scheme to optimize the registration performance. Appropriate selection of a distance or similarity metric is critical to the performance of an image registration method. Three commonly used similarity metrics are introduced in Section 5.1. However, a successful breast MRI registration technique must take into consideration that the image intensity changes over time. This feature of breast MRI means that SSD is not a viable similarity metric even though the images to be registered are from a single modality. CC and MI are both very popular similarity metrics and could be used in our application. In our study, we conduct comparative assessment of the similarity metric in registering DCE breast MRI data.

The overall motion of the breast consists of two forms of motion: global motion, such as that caused by body movements; and the local motion, such as that caused by breathing and the heart beating. Global motion can be characterized using a rigid motion model; while local motion is more appropriately modelled by a non-rigid model.

In our study, we design experiments to assess the performance of similarity metrics in both rigid and non-rigid registration. Since the MR images are always compromised by Rician noise, it is worthy of evaluating the robustness of the similarity metric in the presence of noise. The complicating factors of Rician noise, and non-rigid motion make the problem difficult to analyze in theory and suggest that a numerical study is more appropriate.

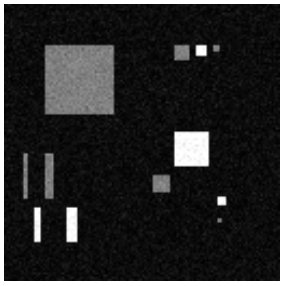
5.3.2 Assessment on Global Motion Correction

A rigid transform can include one or more transformations, such as rotation, translation, and scaling. In our application, the object sizes on image do not vary due to the hardware setup. Therefore we evaluate the performance of similarity metric only on two transformations: rotation and shift.

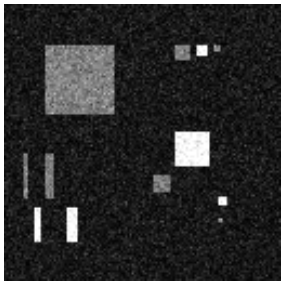
First we computed the similarity measures as a function of the rotation parameter– rotation angle θ. In the experiments, we have a pair of synthetic images(see Fig. 5.2), in which different sizes of objects are placed and the image intensities of the objects change across image pairs. Both CC and MI are maximized at a true value $\theta = 0$ if the images are free of noise. To test the robustness of the similarity metrics, we added noise on the template image (see Fig. 5.2 for two noisy images). Note that the noise is in Rician distribution (noise of both real and imaginary channels are in Gaussian distribution). The performance of similarity metrics in the presence of noise is plotted in Fig. 5.3. We can see that MI outperforms CC with noise at all noise levels.

(a) Reference Image

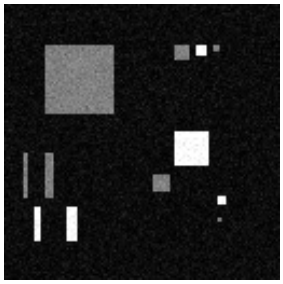
(b)Template Image



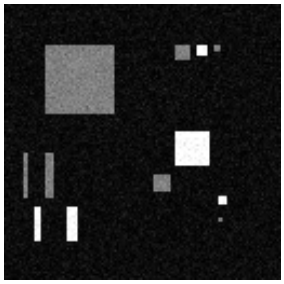
(c) Noisy template1 (12.73%)



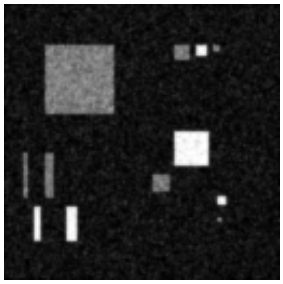
(d)Noisy template2 (28.28%)



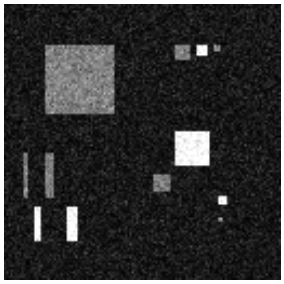
(e) Registered (c) using CC



(f) Registered (c) using MI

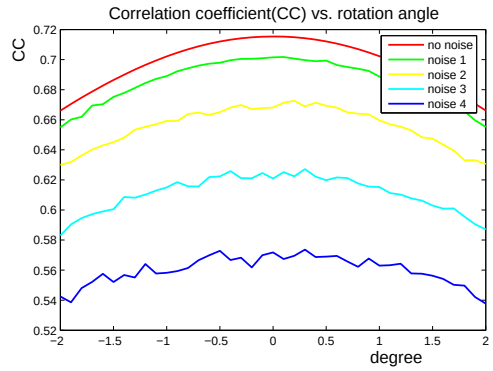


(g) Registered (d) using CC

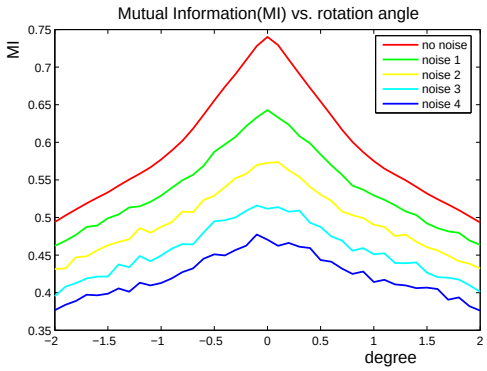


(h)Registered (d) using MI

Figure 5.2: A pair of images simulating different size objects, and the template image is added with low and middle level Rician noise.



(a) CC vs. rotation angle



(b) MI vs. rotation angle

Figure 5.3: Both CC and MI are maximized at a true value $\theta = 0$ if images are free of noise. The noise levels for noise 1, 2, 3, 4 are $6.02\%, 11.3\%, 18.0\%, 22.8\%$ respectively. With noise, CC coefficients are maximized at 0.1, 0.2, 0.3, 0.3 degrees respectively; while MI coefficients are maximized at 0, 0.1, -0.1, -0.1 degrees respectively.

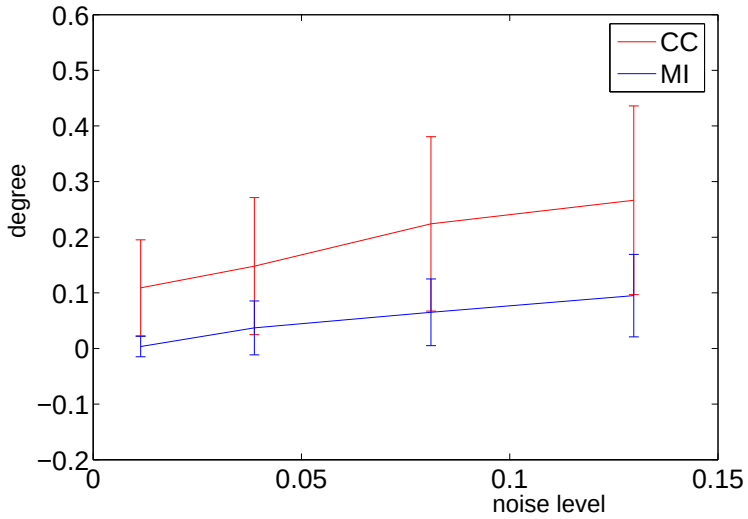


Figure 5.4: Monte Carlo simulation for rotation angle estimate error using similarity metrics MI and CC versus Rician noises

.

The results in Fig. 5.3 were for a single noise realization, so they do not fully characterize the statistics of rotation parameter estimate. We then performed 200 noise realizations at each of these noise levels and computed the rotation parameter estimates $\hat{\theta}$. The mean value and standard deviation of the estimate error are plotted in Fig. 5.4. The MI can still detect the true rotation angle at low level noise, while CC is more vulnerable to noise. The experimental results suggest that MI is more robust than CC for rotation transform.

More relevant here is not the absolute accuracy of the rotation angle, but the resulting image quality after registration. Of particular note is the algorithmic capability in preserving the small sized lesion. In this experiment, different sized objects simulating lesions were placed in the images. Therefore we can assess the registration quality by observing the smallest-size object preserved in the image after registration. See (e,f,g,h) of Fig. 5.2 for four examples.

Similar to the experiments on rotation transform, we comparatively assess the CC and MI regarding the translate transform. The results suggest that MI is more robust than CC.

5.3.3 Assessment on Local Motion Correction

The local deformation of the breast is more appropriately modelled by a non-rigid motion. Several non-rigid models, such as the elastic model and the fluid

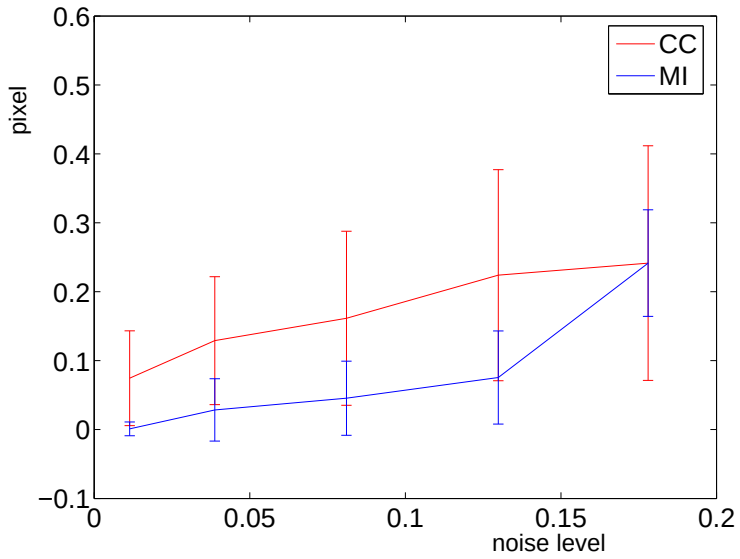


Figure 5.5: Monte Carlo simulation for the translate estimate error using similarity metrics MI and CC versus Rician noise levels

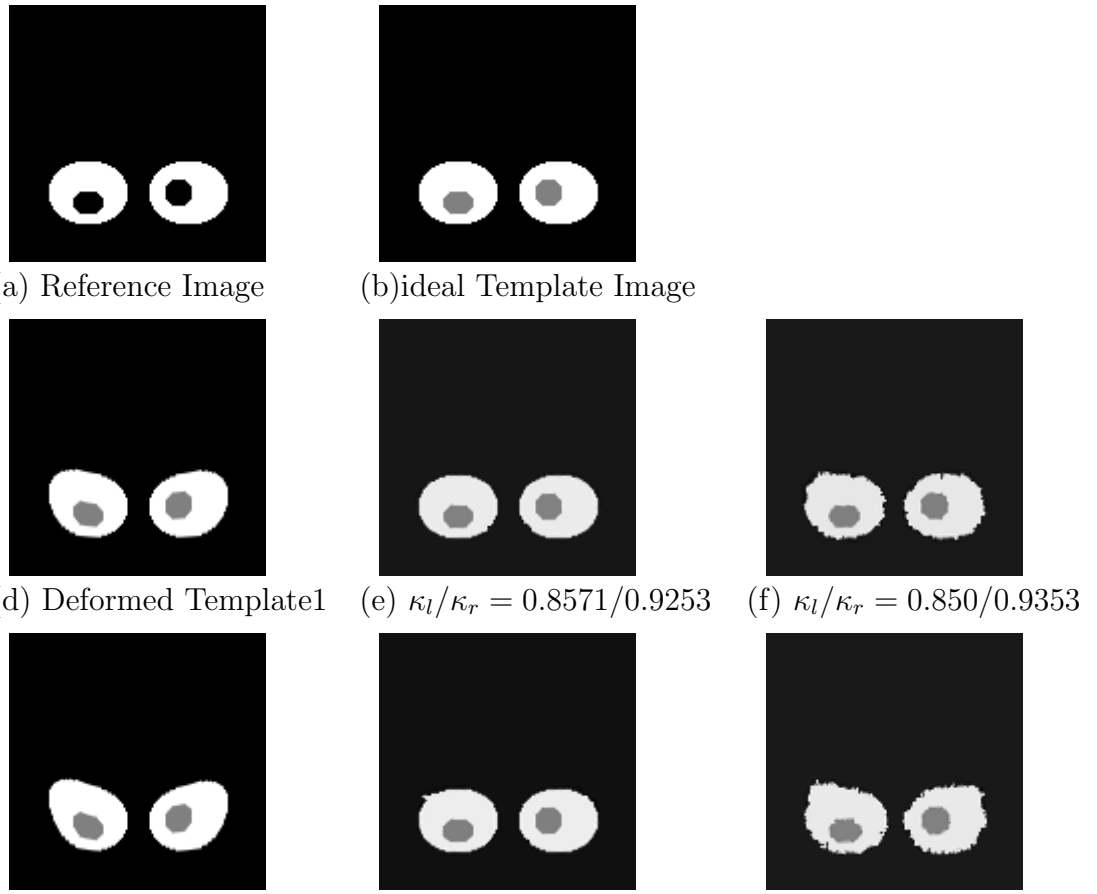
.

model *etc.*, can be used in modelling the local motion. However, the effect of motion model is beyond the scope of this dissertation. We focus on assessing the performance of similarity metrics, CC and MI. In our experiments, we employ the elasticity model in [5]. The problem is to find a spatial displacement field $\mathbf{u}$ that satisfies the following partial differential equation

$$\mu\nabla^2\mathbf{u} + (\lambda + \mu)\nabla(\nabla \cdot \mathbf{u}) = \mathbf{F}(\mathbf{u}) \tag{5.7}$$

where μ and λ are two constants defining the elastic properties of the breast tissue. $\mathbf{F}(\mathbf{u})$ is the variational gradient of the chosen similarity coefficient.

In the synthetic image test cases (See Fig. 5.6), 2 light regions simulating the breasts are subject to shape changes only, while an embedded region in each breast simulating the lesion that takes up contrast is subject to changes in both shape and intensity. The elastic model of 5.7 is applied for motion correction. In term of resulting image quality, more relevant is the accuracy in preserving the objects of interest than the absolute accuracy of the displacement field $\mathbf{u}$. In our study, the object of interest refers to the lesion in the breast. We can compare the level of overlap between the lesion after registration and that in the ground truth image. The kappa metric (κ), also called Dice coefficient, is used for this purpose. The results in Fig. 5.6 are without noise. To test the robustness of the similarity metric in the non-rigid registration, we then performed 200 noise realizations at different noise levels and computed the resulting overlapping metric κ (see Figures 5.7 and 5.8). The numerical results



(a) Reference Image (b)ideal Template Image

(d) Deformed Template1 (e) $\kappa_l/\kappa_r = 0.8571/0.9253$ (f) $\kappa_l/\kappa_r = 0.850/0.9353$

(g) Deformed Template2 (h) $\kappa_l/\kappa_r = 0.7576/0.8943$ (i) $\kappa_l/\kappa_r = 0.7483/0.8989$

Figure 5.6: Image (a) is the reference image, while image (b) is the ground truth template image without motion. In rows 2 and 3, the images of the second column are motion-corrected images of those in the first column employing the CC as the similarity metric; and the images of the third column are motion-corrected images of those in the first column employing the MI as the similarity metric. κ_l and κ_r denote the overlapping coefficient of the registered lesion map and its corresponding ground truth lesion map in the left breast and in the right breast respectively.

suggest that the CC and the MI perform similarly for image with low or middle level noise, and the MI slightly outperforms the CC in the presence of high level noise.

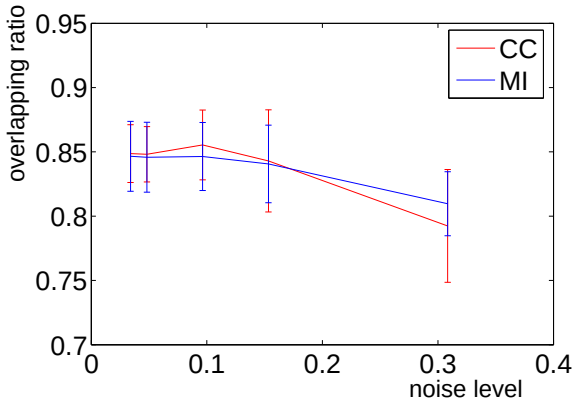


Figure 5.7: Monte Carlo simulation for the overlapping coefficient of the left lesion versus Rician noise

.

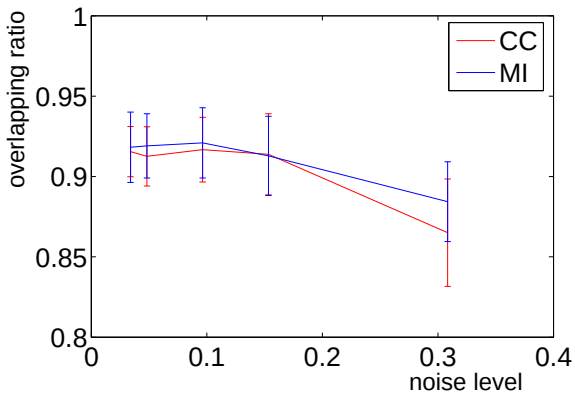


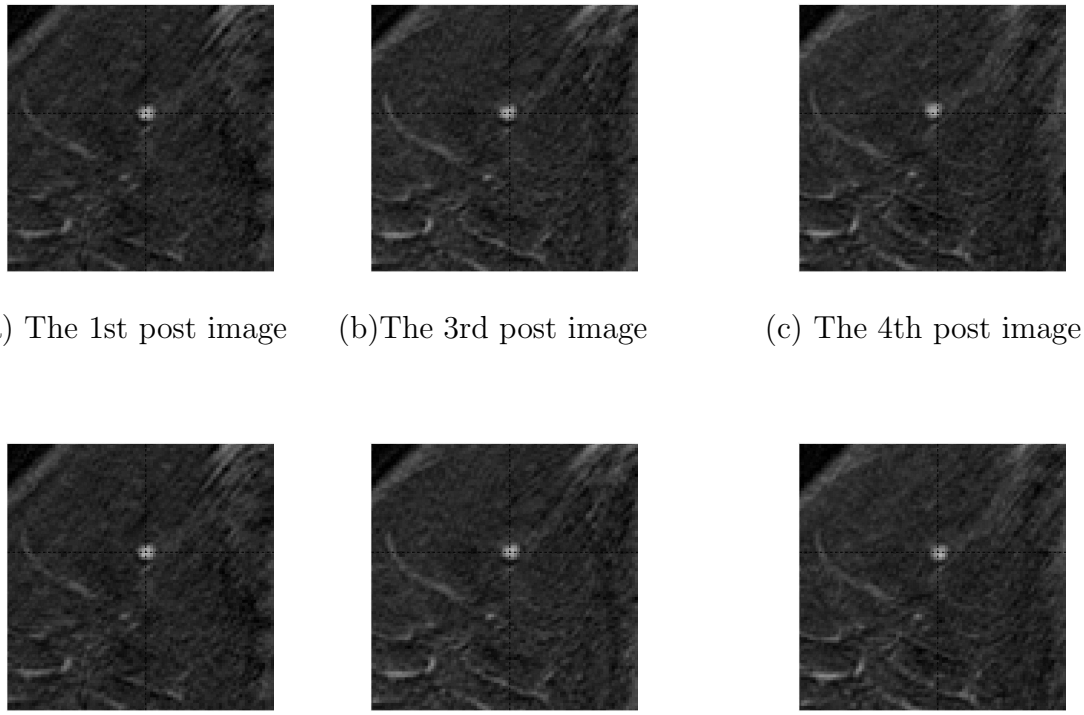
Figure 5.8: Monte Carlo simulation for the overlapping coefficient of the right lesion versus Rician noise

.

5.4 Summary and Registering Clinical DCE Breast MRI data

In this chapter, we comparatively study the performance of the cross correlation (CC) coefficient, and the mutual information (MI) in both rigid and non-rigid registration schemes. We numerically study the impact of similarity coefficient on image quality after registration. The numerical simulation results demonstrated that the performance of the CC is comparable to that of the MI in the presence of low-level noise, and MI outperforms CC for images with middle or high level noise.

In real clinical breast MRI acquisition, the images are susceptible to noise corruption. In our subsequent experiments of registering clinical DCE breast MRI data, we choose MI as the similarity metric and demonstrate its performance with an example shown in Fig. 5.9. The cross-hairs in images indicate the same coordinates in the images. We can observe that there is obvious motion between the 3rd post-contrast image and the 4th post-contrast image compared with the 1st post contrast image (baseline image). Due to motion effect, the kinetic curve directly extracted from raw images doesn't characterize the kinetic features of this benign cyst. After motion correction, the extracted kinetic curve fits more to a typical benign kinetic curve. The kinetic curves before and after registration are shown in Fig. 5.10.



(a) The 1st post image (b)The 3rd post image (c) The 4th post image

(d) The 1st post image (e)Registered 3rd post image (f) Registered 4th post image

Figure 5.9: The top three images (a)-(c) (sagittal view) are raw images acquired from a 60-year-old woman with a history of atypical ductal hyperplasia. The region centered at the cross-hair is a benign cyst at 3 o'clock region of left breast. The 1st post-contrast image is used as baseline image in registration and motion-corrected images are shown in (e) and (f).

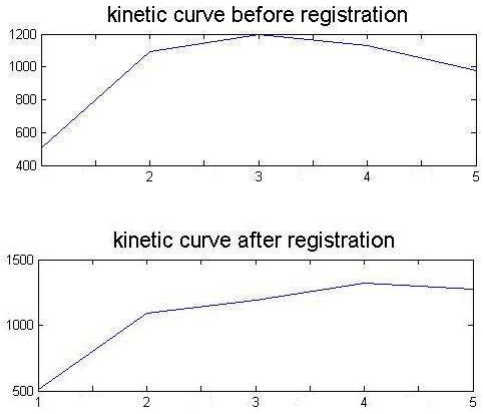


Figure 5.10: Five-point kinetic curves of images in Fig. 5.9. Of each curve, the 1st point represents pre-contrast signal, while the other four points represent four post-contrast signals respectively.

Chapter 6

Classification

The goal of breast MRI is to non-invasively assess abnormalities. For example, if one can be sufficiently certain that a lesion is benign, then the woman can be spared a biopsy. Unfortunately, one of the main criticisms of breast MRI has been its low specificity [89]. Moreover, the specificity of breast MR appears to depend heavily on the radiologist's level of experience [61]. Currently, there is no standardized algorithm for lesion classification. Although it is widely accepted that both morphological and kinetic features on breast MRI should be evaluated in making diagnostic and treatment decisions [60, 79, 95], manual rating of these features is subject to inter- and intra-observer variability [57, 99, 108]. Ideally, the extraction of lesion descriptions could be automated using image processing methods, and then used as inputs for a classifier that outputs a decision on whether the lesion is benign or not.

This chapter presents a classification system based upon quantitative morphological and kinetic features in improving the specificity of breast MRI. Morphological and kinetic features of the lesion are extracted automatically, and then the feature selection step is utilized to select the most relevant features

to optimize the classifier performance. In our study, the area under receiver operating characteristic curve (AUC) is used as the performance metric of the classifier. Results of each stage of the classification system are reported.

6.1 Overview of Previous Algorithms for Lesion Classification

The American College of Radiology has defined a lexicon (Breast Imaging, Reporting, and Data System - BI-RADS [79]) that provides a set of diagnostic features that could be used for developing an automatic classifier. However, current radiological practice involves manual rating of these features, which is subject to inter- and intra-observer variability [57, 99, 108]. Classification based on BI-RADS descriptors [6, 36, 102] and automatic extraction of BI-RADS descriptors [82, 92] have been investigated for other modalities. Some efforts have already been made to automatically predict lesion pathology based on morphological features, kinetic enhancement features, or both [22, 42, 44, 67, 94, 100, 104]. Some such studies [22, 44, 94] have demonstrated that a combination of morphological features and kinetic enhancement features provides better classification than either feature set alone. A previous study also suggests that better classification is obtained when 3D features of the lesion are used as compared to when 2D features are measured slice by slice [44]. As is the case for many classification tasks, previous studies comparing univariate and multivariate regression models for predicting lesion pathology from breast MRI findings have demonstrated that features with marginal clas-

sification performance individually may still yield high predictive values in combination [94]. Several investigations of automatic classification of lesions on DCE breast MRI as benign or malignant are summarized in Table 6.1.

6.2 Our Framework for Lesion Classification

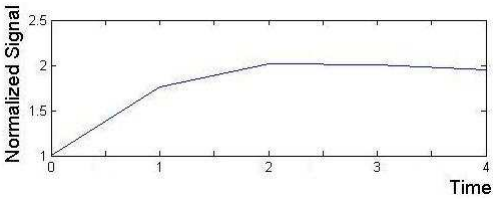
Our long term goal is to automate the workflow of interpreting breast MRI exams and to minimize the interpretation error and observer variability. The goal of the proposed classification study is to develop a model based on morphological and kinetic features that will be extracted automatically and characterized quantitatively using image processing methods. We hypothesize that our proposed segmentation and registration algorithms will yield more accurate delineations of these features and so lead to a more accurate classification performance. There are four major design considerations in developing a classification model: (1) feature extraction, (2) feature selection, (3) classifier training, and (4) evaluation metric. We consider each of these issues in turn in the following paragraphs.

6.2.1 Feature extraction

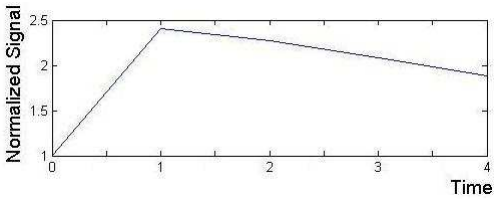
In our study, both the morphological and the kinetic features of the lesion are to be extracted automatically and characterized quantitatively using image processing methods. It should be noted that all kinetic features will be extracted from the motion corrected image series.

Author	Segment	Feature Extraction	Feature Selection	Features	Classifier	Total Cases (Training/Testing)	Classifier Performance
Arbach [4]	SA	A	N/A	Morp	ANN	35	A_z=0.913
Chen [22]	SA	A	Stepwise LDA	Morph, Kin	LDA	121 (LOOCV)	A_z=0.86
Gibbs [42]	M	A	Backward	Kine	LR	43 (no cross-validation)	A_z=0.92
Gilhujs [44]	SA	A	Stepwise LDA	Morph, kine	LDA	27 (LOOCV)	A_z = 0.96
Kinkel [57]	M	M	CART	Morph, kinetic	CART	57 (cross-validation)	Sens.=91% Spec.=83%
Lucht [67]	M	A	N/A	Kine	ANN	258/111	Sens.=84% Spec.=81%
Nunes [75]	M	M	N/A	Morp		98/94	Sens.=96% Spec.=79%
Schnall [94]	M	M	Regression	Morph, kin, clin	Multivariate	995	A_z=0.88
Szabo [100]	M	A	ARD	Morph, kine	ANN	59/46	A_z=0.77
Szabo [101]	X	X	X	Morph, kine		109(79/30)	A_z=0.88
Vomweg [104]	M	M	Genetic Algorithm	Morph, kine,clin	ANN	473/131	Sens.=94% Spec.=92%

Table 6.1: Summary of studies on developing classifiers to predict lesion pathology on breast MRI. A - automated; SA - semi-automated; M - manual; ANN - artificial neural network; LDA - linear discriminant analysis; LR - logistic regression; CART - classification and regression tree; ARD - automatic relevance determination; LOOCV-leave-one-out cross validation. A_z - Area under the ROC.



(a) Mean Curve of Benign Cases



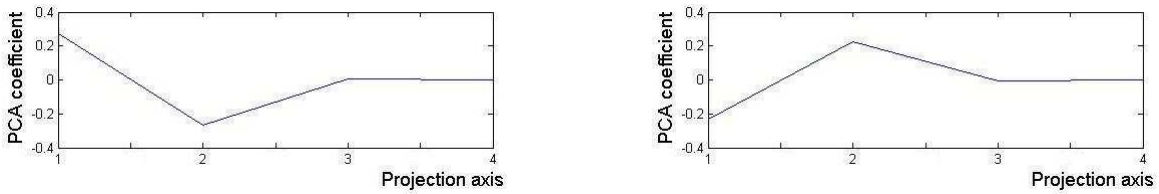
(b)Mean Curve of Malignant Cases

Figure 6.1: Mean kinetic curves of benign cases and malignant cases. The kinetic curves are normalized with respect to the pre-contrast signal of each case and then averaged.

6.2.1.1 Kinetic Features Extraction

In our experiments, the baseline (pre-contrast signal) is normalized as 1; and the post-contrast signals are normalized with respect to their baseline signal. The typical normalized kinetic curves are plotted in Fig. 6.1. Since the baseline signals are always 1 for all lesions (benign and malignant), they do not provide any discriminatory power. From Fig. 6.1, we can observe that the 1st , 2nd and 3rd post-contrast signals together provide rich discriminatory information, and the 4th post-contrast signal may provide additional information. This observation is coincident with conclusions of previous studies that the wash-in rate and the washout-rate are important features to characterize lesion [62].

To extract more salient information from post-contrast signals, a linear feature extraction was obtained by using a linear transform-principle component



(a) Mean PCA coefficients of Benign Cases (b)Mean PCA coefficients of Malignant Case

Figure 6.2: Mean PCA coefficients of benign cases and malignant cases, note that only normalized post-contrast signals are projected into each principal axis and then averaged.

analysis (PCA). PCA coefficients are data formed by projecting the normalized post-contrast signals into the space of the principal components. Therefore we could use the PCA coefficients as a new set of kinetic features. See Fig. 6.2 for mean PCA coefficients of malignant and benign cases respectively. We can observe that the first two major components provide powerful discriminatory information, and the last two components are of little use in discriminating the benign case from the malignant one.

6.2.1.2 Morphological Features Extraction

After the lesion is segmented out from the $3D$ post-contrast image, some geometric features of the lesion can be extracted, such as its *spiculation*, *aspect-ratio* and *compactness*.

Spiculation is computed as the standard deviation of the radial length of all points on the segmented lesion surface. The radial length is defined as the

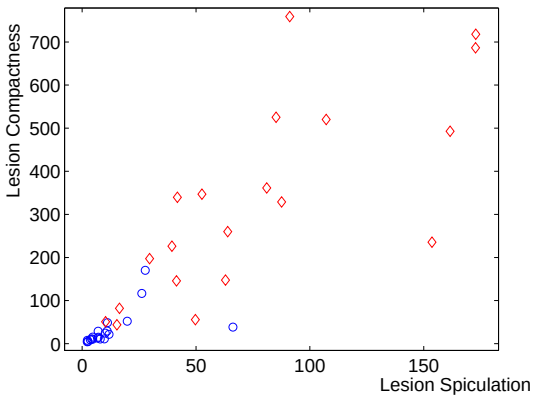


Figure 6.3: Distribution of morphological features sampled from 40 cases (20 benign cases and 20 malignant cases). The blue circles denote benign cases while the red diamonds denote malignant lesions.

distance from a boundary point to the geometric center of the lesion. The *compactness* of a mass lesion can be defined as the ratio of its surface area to its volume. Ideally, the compactness of a round or an oval mass is small, while that of a spiculated mass is large. Thus, the *compactness* of a mass is another quantitative measure of the likelihood of malignancy. The *aspect-ratio* can be defined as ratio of the *maximal diameter* and to the *minimal diameter* of the bounding box, and this ratio is a quantitative measure of the likelihood that the mass is malignant [20]. Ideally, the *aspect-ratio* of round masses is close to 1, that of oval masses is larger than 1 and that of lobular masses is even larger.

The aforementioned morphological features have been efficiently utilized in previous work to classify breast lesions [4, 55]. In Fig. 6.3, we plot a distribution map of spiculation and compactness.

The morphological features, such as *spiculation* and *compactness*, can be extracted from both the 3D lesion and from the 2D representative slice. In our experiments, we extracted the 3D and 2D morphological features and used the feature selection step to determine the most relevant features for classifiers.

6.2.2 SVM-based Classifier

A variety of statistical machine learning algorithms could be used to classify a suspicious lesion as likely benign or likely malignant based on the extracted features. The classifiers, such as linear discriminant analysis (LDA) [32] and support vector machine (SVM) [15, 29] could be utilized for our classification purpose. SVMs are a set of machine learning methods used for classification and have come into widespread use for many biomedical classification tasks. We discuss the principles of SVM in the following paragraphs.

Suppose we have a set of data points $D = \{(\mathbf{x}_i, y_i)|\mathbf{x}_i \in \Re^p, y_i \in \{-1, 1\}\}_{i=1}^n$. We can construct a separating hyperplane that divides those points having $y_i = 1$ from those having $y_i = -1$, such a hyperplane can be denoted as

$$\mathbf{x} : f(\mathbf{x}) = \mathbf{x}^T\beta + \beta_0 = 0 \tag{6.1}$$

where β is an unit normal vector perpendicular to the hyperplane, see Fig. 6.4. Our goal is to maximize the margin (distance) between data of two classes, and this problem can be formulated as

$$max\ \mathrm{C}\ \ \text{s.t.}\ y_i(x_i^T\beta + \beta_0) \geq C, i = 1, \ldots, n \tag{6.2}$$

The optimal hyperplane can be determined by using only those data points on the margin, so called *support vector*s [29]. The optimization problem 6.2 can be reformulated as

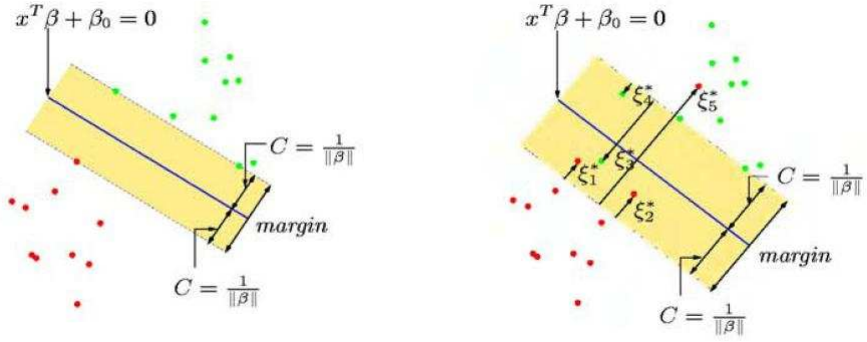
$$min\|\beta\| \text{ s.t. } y_i(x_i^T\beta + \beta_0) \geq C, i = 1, \ldots, n \tag{6.3}$$

Very often, the data cannot be cleanly separated by a hyperplane, thus there are some mislabeled data. The method by Cortes and Vladimir [29] introduced slack variables, ξ_i, to measure the amount of misclassification of $\mathbf{x}_i$. Therefore, the optimization problem for non-separable case can be formulated as

$$min\|\beta\| \text{ s.t. } \begin{cases} y_i(x_i^T\beta + \beta_0) \geq 1 - \xi_i, \forall i \\ \xi_i \geq 0, \sum \xi_i \leq K \end{cases} \tag{6.4}$$

where K is a constant. By bounding the total misclassification amount $\sum \xi_i$, we bound the total number of mislabeled data.

In order to extend the SVM methodology to handle data that is not linearly separable, a non-linear kernel function was suggested to map the data into a higher dimensional space that is linearly separable. Commonly used non-linear kernel functions include the radial basis function (RBF) and the Sigmoidal kernel. The SVM will be applied to solve our classification problem in this dissertation project. Since it cannot be known *a prior* if benign and malignant lesions are linearly separable in the proposed feature space, we need to compare the performance of the linear kernel to that of the non-linear kernel when applying SVM. The libsvm software package [19] was used to implement the



(a) Separable case (b) Non-separable case

Figure 6.4: Support vector classifiers. The red and green points represent data belonging to different classes. In (a), the maximal margin width is $2C$. In (b), the points with ξ_i^* represents those mislabeled data on the wrong side of margin by an amount of $C\xi_i$; and points corrected labelled have $\xi_i^* = 0$. Picture courtesy of [37].

SVM classifier and the RBF kernel was selected as the non-linear kernel in our experiments.

6.2.3 Feature Selection and Ranking

Some features may prove to be irrelevant or redundant, which can degrade the classifier performance [47]. Feature selection refers to the process of identifying a subset of features that is most valuable for classification according to quantitative criterion [47].

In order to accomplish optimal classification performance, each variable or feature needs to be evaluated on its effect on the classification to identify relevant

features. Of the existing feature ranking methods, some are related to certain classification methods while some are independent of classifiers. For example, the Fisher-score, a correlation coefficient between the feature and the label, is a simple and efficient criterion in ranking features. In addition to Fisher-score, we also evaluated subset relevance based on the SVM criteria, such as the weight vector norm and upperbounds of the leave-one-out error, with respect to a variable [88]. The latter strategy has the advantage of yielding high classification performance, as confirmed by the experimental results. Our feature ranking procedure is implemented using the software package downloaded from webpage of the first author of [88].

6.2.4 Classifier Training and Evaluation Metric

The ideal classifier training process requires three independent sets: a training set, a validation set, and a testing set. However, when the available data are limited, as is the case in this pilot study, a cross-validation strategy is used. In K-fold cross validation, the data set is randomly partitioned into K disjointed sets of equal size. The classifier is trained K times with a different set held out as the validation set each time. The special case of K-fold cross-validation in which K is equal to the number of samples is referred as leave-one-out cross-validation or round-robin training. Leave-one-out cross-validation has been widely adopted for the analysis of biomedical classification tasks based on small data sets and was also employed in this dissertation project.

As it is the standard practice in evaluating clinical decision aids for binary classification tasks, receiver operating characteristic (ROC) analysis will be used for evaluating the classifier developed in our study. An ROC curve shows the tradeoffs in sensitivity and specificity that can be achieved by varying the threshold on the decision variable [76]. The area under the ROC curve (AUC) is a common summary measure of the performance of a classifier, and standard statistical methods exist for comparing the AUCs of different models or feature sets.

6.3 Classification Experimental Results

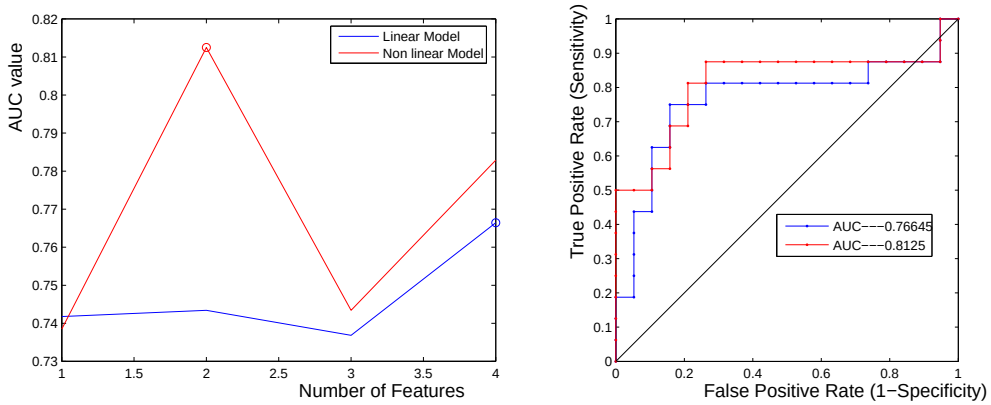
In this section, we describe our experiments in evaluating the extracted morphological and kinetic features. We first performed classifications on kinetic and morphological features independently, and selected the features to optimize the classification performance respectively. The features were compared using the feature selection method and the leave-one-out classification procedure explained in Section 6.2. We then combined the selected features, and repeated the feature selection step on the combined feature set in order to select features that yielded the best overall classification performance. The experimental results are detailed in following paragraphs.

6.3.1 Classification Experiment on Kinetic Features

We can do lesion classification on the normalized post-contrast signals or extracted PCA coefficients. To compare these two feature sets, we applied the same feature selection and classification procedure on each of them separately.

We first used the normalized post-contrast signals for classification. There are signals at four post-contrast time points. We ranked them according to the method described in Section 6.2.3. See Fig.6.5 for the ROC performance vs. the number of features used. The ROC curve for best AUC values in Fig.6.5(a) are plotted in Fig.6.5(b). For linear model, all post-contrast signals are needed to maximize the classification performance. For nonlinear model, the first post-contrast value and the late post-contrast are sufficient to produce the best result. By using the normalized post-contrast signals as features, the nonlinear SVM model resulted in a better performance ($A_z = 0.813$) than the linear SVM model ($A_z = 0.766$).

We then use the PCA coefficients as classifier inputs. The PCA coefficients are data formed by projecting the normalized post-contrast signals into the space of the principal components. We can use the PCA coefficients as a new set of kinetic features. See Fig.6.6 for the ROC performance versus the number of features used. The ROC curve for best AUC values in Fig. 6.6(a) are plotted in Fig. 6.6(b). For the linear model, only two major components are needed to maximize the classification performance. For the nonlinear model,



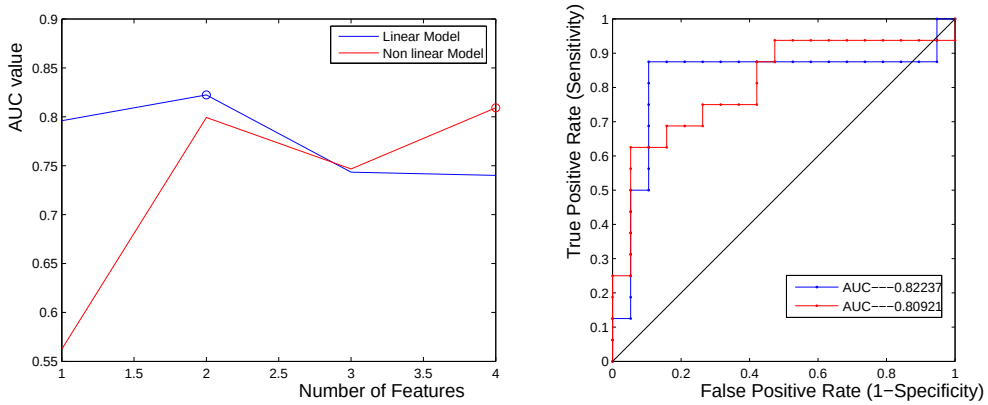
(a) AUC *vs.* ranked post-contrast signals (b)ROC curves

Figure 6.5: Classification results using normalized post-contrast signals as kinetic features.

all PCA components are needed to produce the best result. By using the PCA coefficients as classifier inputs, the linear SVM model resulted in a better performance of ($A_z = 0.822$) than the non-linear SVM model ($A_z = 0.809$).

6.3.2 Classification Experiment on Morphological Features

Two steps were performed to evaluate the segmentation algorithms and their derived features. First, we selected the most relevant features that maximize the classification accuracy. The experiments were tested on a set of features extracted from manually segmented lesions. We tested the morphological features using both the linear kernel SVM and the non-linear (RBF) kernel SVM methods. The features were ranked by the method described in [88], and the results are presented in Fig. 6.7. We can see that the non-linear kernel



(a) AUC *vs.* ranked PCA coefficients (b)ROC curves

Figure 6.6: Classification results using PCA coefficients as kinetic features.

method outperforms the linear kernel method for classification when using only the morphological features of lesion, and the compactness index of the representative 2D slice and 3D spiculation index are the two most relevant features that optimize the classifier performance in using non-linear kernel SVM.

The second step was to evaluate the impact of different segmentation algorithms on lesion classification. According to classification results based on manually segmented lesions, the two most relevant morphological features are the compactness index of representative 2D slice and 3D spiculation index. We then extracted these two morphological features from lesions segmented by different algorithms, such as threshold-based method, mean-shift method, Fuzzy c-means method, and the STAPLE method. SVM methods were employed to compute classify lesions based on their morphological features. The classification results are summarized in Table 6.2. Based on the algorithmic segmentation outputs of 40 lesions (20 benign and 20 malignant), the STAPLE

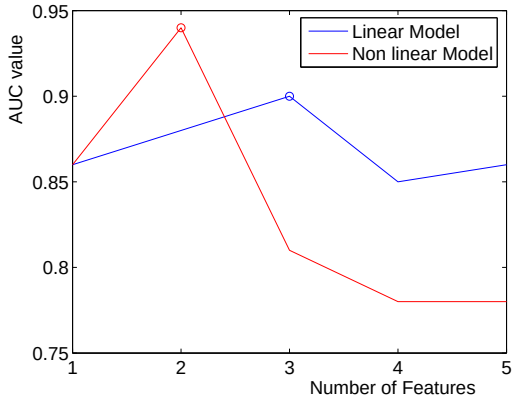
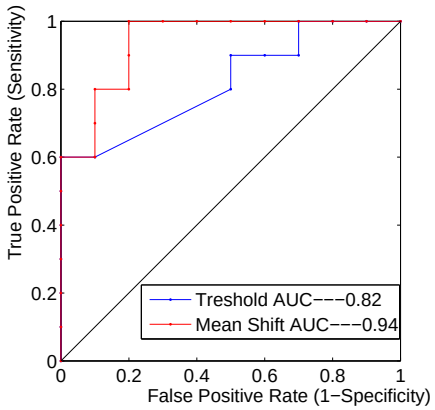
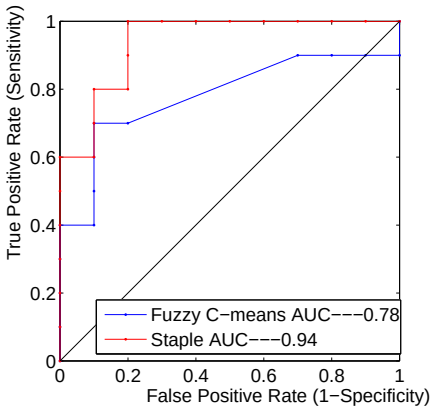


Figure 6.7: AUC *vs.* number of morphological features using different SVM kernels.



(a) Threshold and mean shift methods



(b) Fuzzy C-means and STAPLE methods

Figure 6.8: Classification results using morphological features from lesions segmented out by different algorithms.

Segmentation method	Accuracy	AUC
Volume Threshold	15/20	0.82
Fuzzy C-means	13/20	0.78
Mean Shift	17/20	0.94
Staple	17/20	0.94

Table 6.2: Best classification performance metrics of lesion segmented out by different algorithms

and mean shift algorithms yielded better classification results than the Fuzzy C-means and the threshold-based methods.

6.3.3 Classification Experiment on Combined Features

According to results of previous experiments, we know that the most relevant morphological features are the compactness index of representative 2D slice and 3D spiculation index, and the most relevant kinetic features are the first two major PCA coefficients of normalized post-contrast signals. In this experiment, we performed classification using a combination of the most relevant morphological features and most relevant kinetic features, and this new set of features were again applied using feature selection procedure and 5-fold cross-validation procedure described in previous section. The best classification results are summarized in Table 6.3.

From Table 6.3, we can observe that combining morphological and kinetic features improves the AUC value or the best classification rate as compared to using the morphological features or the kinetic features alone. The experimental results are consistent with findings of previous studies that both

Feature	AUC	Selected features
Morphological	0.94	spiculation(3D) and compactness(2D)
Kinetic	0.822	First two major PCA coefficients
Morphological + Kinetic	0.952	spiculation(3D), compactness(2D), and first two PCA coefficients

Table 6.3: Best classification performance and their corresponding sets of features used in lesion classification

morphological and kinetic features are important in distinguishing between malignant and benign breast lesions.

Chapter 7

Summary and Future Directions

7.1 Summary of this dissertation

This dissertation is a pilot study of developing a computational image analysis prototype system for an increasingly used breast imaging modality–dynamic contrast-enhanced breast MRI. We developed image analysis methods that ultimately may improve the performance of each of the major algorithm components that would be needed in a breast MRI CADx system.

In realizing our prototype system, we utilize both morphological and kinetic features of the lesion on breast MR images. To extract morphological features, the lesion needs to segmented out from its surrounding tissues; we utilized STAPLE algorithm to combine segmentation maps generated by several algorithms, and further evaluate performance of segmentation algorithms by their impact on subsequent classification system. To compensate for the motion across image volumes acquired sequentially, we comparatively assessed the similarity metrics, cross correlation coefficient and mutual information, in registering DCE breast MR images. Based upon quantitative morphological and kinetic features, we implemented a classification system in order to

improve the specificity of breast MRI. The morphological and the kinetic features of the lesion were extracted automatically, and then a feature selection step was utilized to select the most relevant features to maximize the classifier performance. In our study, the area under ROC curve (AUC) is used as the performance metric of the classifier, and our results are competitive with those of previous studies.

As breast MR imaging hardware equipment and acquisition techniques are evolving, for example, high field scanners, such as 3T, have been shown to produce more clearer delineation of tumor; and fast imaging methods are still being developed to accelerate image acquisition; however, we expect that the image analysis methods developed in this project can be similarly applied on images acquired by future techniques or scanners.

7.2 Future Directions

In the future, the breast MRI will be increasingly and routinely used, as a diagnostic, biopsy, staging, and treatment monitoring technique in breast cancer patient care. Research on breast image analysis is essential to fully exploit and efficiently utilize the imaging information that has been acquired. Here we discuss two aspects that are critical in developing future breast image analysis systems.

7.2.1 Multimodality Imaging Information Integration

In clinical practice, very often the information from multiple modalities instead of just one modality is used in making a diagnosis. The information from different modality can complement to each other, for example, mammography-occulted cancers may be detected by breast MRI or ultrasound. Also the information from different modality can provide correlative findings, for example, breast MRI can be used to further evaluate a finding on mammography or breast ultrasound. Therefore it is necessary to develop a breast image analysis system that is capable of integrating multimodality information.

A challenge in imaging information integration involves image fusion/registration that would establish the correspondence of the same physical object on images from different modalities. This involves registration from 2D images (mammogram or ultrasound image) to 3D images (tomosynthesis, or breast CT, or breast MRI), from 2D images (mammogram) to 2D images (breast ultrasound), and from 3D images (breast MRI) to 3D images (breast CT or tomosynthesis).

7.2.2 Breast Image Database

One should recognize that the imaging scanners from different manufacturers and the image acquisition protocols at different sites will be variable. The comparison and evaluation of image analysis techniques against each other requires common data sets for evaluation, which is a starting point for developing clinically-applicable image analysis methods. Additionally, researchers

working on breast image analysis usually have access to a small number of single modality or multi-modality breast images on which to develop and test their work. For the purpose of promoting breast image analysis, there is a critical need to construct a database of breast images as research resources.

The proposed database should contain several image data sets. For example, X-ray mammography data, ultrasound images, and breast MR images should be included. All abnormalities on images should be confirmed by pathology reviews. There also should be a consensus criteria for preparation and submission of cases that are representative of clinical practice to include typical types of tumors. Such a database can be used not only as a testbed for image analysis methods, but also as a teaching and training resource for both human readers and CAD.

Bibliography

[1] http://www.ismrm.org/special/NSF.htm.

[2] 2007. FDA Requests Boxed Warnings for Contrast Agents Used to Improve MRI Images.

[3] Mohamed N. Ahmed, Sameh M. Yamany, Nevin Mohamed, Aly A. Farag, and Thomas Moriarty. A modified fuzzy c-means algorithm for bias field estimation and segmentation of MRI data. *IEEE Transactions on Medical Imaging*, 21(3):193–199, March 2002.

[4] Lina Arbash, Alan H. Stolpen, and Joseph M. Reinhardt. Classification of breast MRI lesions using a backpropagation neural network (BNN). In *Proceedings of IEEE International Sympoisum on Biomedical Imaging*, pages 253–256, 2004.

[5] Ruzena Bajcsy and Stane Kovačič. Multi-resolution elastic matching. *Computer Vision, Graphics and Image Processing*, 46:1–21, 1989.

[6] Jay A. Baker, Phyllis J. Kornguth, Joseph Y. Lo, and Carey E. Floyd, Jr. Artificial neural network: improving the quality of breast biopsy recommendations. *Radiology*, 198(1):131, 1996.

[7] Wendle A. Berg, Lorena Gutlerrez, Morlel S. NessAiver, W. Carter, Mythreyl Bhargavan, Rebecca S. Lewis, and Olga. B. Loffe. Diagnostic

accuracy of mammography, clinical examination, US, and MR imaging in preoperative assesment of breast cancer. *Radiology*, 233:830–849, 2004.

[8] Richard E. Bird, Terry W. Wallace, and Bonnie C. Yankaskas. Analysis of cancers missed at screening mammography. *Radiology*, 184:613–617, 1992.

[9] Beata Boné, Botond K. Szabó, L.G. Perbeck, B. Veress, and Peter Aspelin. Can contrast-enhanced MR imaging predict survival in breast cancer. *Acta Radiologica*, 44:373–378, 2003.

[10] Georgg Bongartz and Walter Kucharczyk. Nephrogenic systemic fibrosis: Summary of the special symposium. *Journal of Magnetic Resonance Imaging*, 26(5):1179–1181, November 2007.

[11] John M. Boone, Thomas R. Nelson, Karen K. Lindfors, and J. Anthony Seibert. Dedicated breast CT: Radiation dose and image quality evaluation. *Radiology*, 221:657–667, 2001.

[12] Y. Boykov, V. Lee, H. Rusinek, and R. Bansal. Segmentation of dynamic N-D data sets via graph cuts using Markov models. In *Medical Image Computing and Computer-Assisted Intervention MICCAI,LNCS 2208*, pages 1058–1066, 2001.

[13] Morten Bro-Nielsen and Claus Grankow. Fast fluid registration of medical images. In *Visualisation in Biomedical Computing,LNCS 1131*,

pages 267–276, 1996.

[14] Giovanni A. Buonaccorsi, Caleb Roberts, Sue Cheung, Yvonne Watson, Karen Davies, Alan Jackson, Gordon Jayson, and Geoff J.M. Parker. Tracer kinetic model-driven registration for dynamic contrast enhanced MRI time series. In *Medical Image Computing and Computer-Assisted Intervention MICCAI,LNCS 3749*, pages 91–98, 2005.

[15] Christopher J.C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2:121–167, 1998.

[16] Jerrold Bushberg, J. Anthony Seibert, Edwin M. Leidholdt, Jr., and John M. Boone. *The Essential Physics of Medical Imaging.* Lippincott Williams & Wilkins, second edition, 2001.

[17] R. C. and R. E. Woods. *Digital Image Processing.* Prentice Hall, 2nd edition, 2002.

[18] Heang-Ping Chan, Berkman Sahiner, Kwok Leung Lam, Nicholas Petrick, Mark A. Helvie, Mitchell M. Goodsitt, and Dorit D. Adler. Computerized analysis of mammographic microcalcifications in morphological and texture feature spaces. *Medical Physics*, 25(10):2007–2019, 1998.

[19] Chih-Chung Chang and Chih-Jen Li. LIBSVM: a library for support vector machines. Technical report, 2001. Software available at `http:`

//www.csie.ntu.edu.tw/~cjlin/libsvm.

[20] Ruey-Feng Chang, Wen-Jie Wu, Woo Kyung Moon, and Dar-Ren Chen. Automatic ultrasound segmentaion and morphology based diagnosis of solid breast tumors. *Breast Cancer Research and Treatment*, 89(2):179–185, 2005.

[21] Weijie Chen, Maryellen L. Giger, and Ulrich Bick. A fuzzy c-means (FCM)-based approach for computerized segmentation of breast lesions in dynamic contrast-enhanced MR images. *Academic Radiology*, 13(1):63–72, Jan. 2006.

[22] Weijie Chen, Maryellen L. Giger, Li Lan, and Ulrich Bick. Computerized segmentation of breast MRI: Investigation of enhanced-variance dynamics. *Medical Physics*, 31(5):1076–1082, 2004.

[23] Yizong Cheng. Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8):790–799, 1995.

[24] Gary E. Christensen. *Deformable shape models for neuroanatomy.* PhD thesis, Washington University at St. Louis, 1994.

[25] David J. Collins and Anwar R. Padhani. Dynamic magnetic resonance imaging of tumor perfusion: Approaches and biomedical challenges. *IEEE Engineering in Medicine and Biology Magazine*, 23(5):65–83, 2004.

[26] Dorin Comaniciu. *NonParametric Robust Methods for Computer Vision*. PhD thesis, The State University of New Jersey at Rutgers, 2000.

[27] Dorin Comaniciu and Peter Meer. Mean shift analysis and applications. In *Proceedings of IEEE Intl. Conf. on Computer Vision*, pages 1197–1203, 1999.

[28] Dorin Comaniciu and Peter Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619, 2002.

[29] Corinna Cortes and Vladimir Vapnik. Support vector networks. *Machine Learning*, 20:273–297, 1995.

[30] Flavio Crippa, Alberto Gerali, Alessandra Alessi, Roberto Agresti, and Emilio Bombardieri. FDG-PET for axilliary lymph node staging in primary breast cancer. *European Journal of Nuclear Medicine and Molecular Imaging*, 31:S97–S102, 2004.

[31] P. Davis and K. McCarty Jr. Sensitivity of enhanced MRI for the detection of breast cancer: new, mulicentric, residual, and recurrent. *European Radiology*, S5:S289–S298, 1997.

[32] Richard O. Duda, David G. Stork, and Peter E. Hart. *Pattern Classification*. Wiley-Interscience, second edition, 2001.

[33] Laura Esserman, Nola Hylton, Leila Yassa, John Barclay, Steven Frankel, and Edward Sickles. Utility of magnetic resonance imaging in the man-

agement of breast cancer: Evidence for improved preoperative staging. *Journal of Clinical Oncology*, 17(1):110–119, January 1999.

[34] Laura Esserman, Dulcy Wolverton, and Nola Hylton. Magnetic resonance imaging for primary breast cancer management: current role and new applications. *Endocrine-Related Cancer*, 9:141–153, 2002.

[35] Mark Fashing and Carlo Tomasi. Mean shift is a bound optimization. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27(3), 2005.

[36] Carey E. Floyd Jr., Joseph Y. Lo, and Georgia D. Tourassi. Case-based reasoning computer algorithm that uses mammographic findings for breast biopsy decisions. *American Journal of Roentgenology*, 175(5):1347–1352, 2000.

[37] Jerome Frideman, Trevor Hastie, and Robert Tibshirani. *The elements of statistical learning*. Springer, first edition, 2001.

[38] Keinosuke Fukunaga and Larry Hostetler. The estimation of the gradient of a density function, with application in pattern recognition. *IEEE Transactions on Information Theory*, 21:32–40, 1975.

[39] James C. Gee, Martin Reivich, and Ruzena Bajcsy. Elasticially deforming atlas to match anatomical brain images. *Journal of Computer Assistes Tomography*, pages 225–236, 1993.

[40] Borgdan Georgescu, Ilan Shimshoni, and Peter Meer. Mean shift based clustering in high dimensions: A texture classification example. In *Proceedings of IEEE International Conference on Computer Vision*, 2003.

[41] Guido Gerig, Matthieu Jomier, and Miranda Chakos. Valmet: A new validation tool for assessing and improving 3D object segmentation. *MICCAI 2001: Lecture Notes in Computer Science*, 2208:516–523, 2001.

[42] P. Gibbs, G. P. Liney, M. Lowry, P. J. Kneeshaw, and L. W. Turnbull. Differentiation of benign and malignant sub-1 cm breast lesion using dynamic contrast enhanced MRI. *The Breast*, 13(2):115–121, 2004.

[43] Maryellen Giger. Computerized analysis of images in the detection and diagnosis of breast cancer. *Seminars in Ultrasound, CT and MRI*, 25:411–418, 2004.

[44] Kenneth G. Gilhuijs, Maryellen L. Giger, and Ulrich Bick. Computerized analysis of breast lesions in three dimensions using dynamic magnetic-resonance imaging. *Medical Physics*, 25(9):1647–1654, 1998.

[45] Kenneth G. Gilhuijs, Maryellen L. Giger, and Ulrich Bick. A method for computerized assessment of tumor extent in contrast-enhanced mr images of the breast. In K. Doi, H. MacMahon, M.L. Giger, and K.R. Hoffmann, editors, *Computer-Aided Diagnosis in Medical Imaging*, pages 305–310. Elsevier Science, 1999.

[46] Stacy K. Goergen, Jill Evans, Gary P. B. Cohen, and Jamie H. MacMillan. Characteristics of breast carcinomas missed by screening radiologists. *Radiology*, 204:131–135, 1997.

[47] Isabelle Guyon and André Elisseeff. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3:1157–1182, 2003.

[48] Joseph V. Hajnal, Derek L. G. Hill, and David J. Hawkes, editors. *Medical Image Registration*. CRC Press, 2001.

[49] Ferris M. Hall. The rise and impending decline of screening mammography. *Radiology*, 247(3):597–601, June 2008.

[50] Paul Hayton. *Analysis of contrast-enhanced breast MRI*. PhD thesis, Oxford University, 1998.

[51] Paul Hayton, Michael Brady, Lionel Tarassenko, and Niall Moore. Analysis of dynamic MR breast images using a model of contrast enhancement. *Medical Image Analysis*, 1(3):207–224, 1996.

[52] R. Edward Hendrick, editor. *Breast MRI: Fundamentals and Technical Aspects*. Springer, 2007.

[53] David L.G. Hill, Philipp G. Batchelor, Mark Holden, and Derek J. Hawkes. Medical image registration. *Physics in Medicine and Biology*, 46:R1–R45, 2001.

[54] Linda L. Humphrey, Mark Helfand, Benjamin K. Chan, and Steven H. Woolf. Breast cancer screening: A summary of the evidence for the U.S. Preventive Service Task Force. *Annual Internal Medicine*, 137:347–360, 2002.

[55] Z. Huo, M. L. Giger, C. Vyborny, D. Wolverton, R. Schmidt, and K. Doi. Automated computerized classification of malignant and benign mass lesions on digitized mammograms. *Academic Radiology*, 5:155–168, 1998.

[56] Nola Hylton. Magnetic resonance imaging of the breast: Opportunities to improve breast cancer management. *Journal of Clinical Oncology*, 23(8):1678–1684, March 2005.

[57] Karen Kinkel, Thomas H. Helbich, Laura J. Esserman, John Barclay, Ellen H. Schwerin, Edward A. Sickles, and Nola M. Hylton. Dynamic high-spatial-resolution MR imaging of suspicious breast lesions: diagnostic criteria and interoberserver variability. *American Journal of Roentgenology*, 175(1):35–43, 2000.

[58] Daniel B. Kopans. *Breast Imaging*. Lippincottt Williams Wikins, third edition, 2006.

[59] Christiane Katharina Kuhl, 2006. Coursenote 'Problem Solving with Breast MRI: Diagnostic Utility of Breast MR' on the 14th annual conference of International Society of Magnetic Resonance in Medicine(ISMRM).

[60] Christiane Katharina Kuhl, 2006. Coursenote 'MRI Criteria to Diagnose Breast Cancer' on the 14th annual conference of International Society of Magnetic Resonance in Medicine(ISMRM).

[61] Christiane Katharina Kuhl, 2006. Coursenote 'Diagnostic Utility of breast MRI' on the 14th annual conference of International Society of Magnetic Resonance in Medicine(ISMRM).

[62] Christiane Katharina Kuhl, Peter Mielcareck, Sven Klaschik, Claudia Leutner, Eva Wardelmann, Jürgen Gieseke, and Hans H. Schild. Dynamic breast MR imaging: Are signal intensity time course data useful for differential diagnosis of enhancing lesions. *Radiology*, 211:101–110, 1999.

[63] Christiane Katharina Kuhl, Hans H. Schild, and Nuschin Morakkabati. Dynamic bilateral contrast-enhanced MR imaging of the breast: Trade-off between spatial and temporal resolution. *Radiology*, 236:789–800, 2005.

[64] Constance D. Lehman, Jeffrey D. Blume, Paul Weatherall, David Thickman, Nola Hylton, Ellen Warner, Etta Pisano, Stuart J. Schnitt, Constantine Gatsonis, and Mitchell Schnall. Screening women at high risk for breast cancer with mammography and magnetic rsonance imaging. *Cancer*, 103:1898–1905, 2005.

[65] Peter Lind, Isabel Igerc, Thomas Beyer, Peter Reinprecht, and Klaus Hausegger. Current and future use of positron emission tomography

(PET) in breast cancer. *European Journal of Nuclear Medicine and Molecular Imaging*, 31:S125–S134, 2004.

[66] Karen K. Lindfors, John M. Boone, Thomas R. Nelson, Kai Yang, Alexander L. C. Kwan, and Dewitt F. Miller. Snakes, shapes, and gradient vector flow. *Dedicated Breast CT: Initial Clinical Experience*, 246(3):725–733, March 2008.

[67] Robert E. Lucht, Michael V. Knopp, and Gunnar Brix. Classification of signal-time curves from dynamic MR mammography by neural networks. 19(1):51–57, 2001.

[68] David A. Mankoff and William B. Eubank. Current and future use of positron emission tomography (PET) in breast cancer. *Journal of Mammary Gland Biological Neoplasia*, 11:125–136, 2006.

[69] MARIBS study group. Screening with magnetic resonance imaging and mammography of a uk population at high familial risk of breast cancer: a prospective multicentre cohort study (MARIBS). *Lancet*, 365:1769–1778, May 2005.

[70] Tim McInerney and Demetri Terzopoulos. Deformable models in medical image analysis:a survey. *Medical Image Analysis*, 1(2):91–108, 1996.

[71] Lina Arbach Meinel and Joseph M. Reinhardt. Breast MRI computer-aided diagnosis system. In Jasjit S. Suri and Rangaraj M. Rangayyan,

editors, *Recent Advances in Breast Imaging, Mammography, and Computer-Aided Diagnosis of Breast Cancer*, pages 791–832. SPIE Press, 2006.

[72] André E. Merbach and Éva Tóth, editors. *The Chemistry of Contrast Agents in Medical Magnetic Resonance Imaging.* John Wiley and Sons Ltd, 2001.

[73] Keith L. Moore and Arthur F. Dalley. *Clinically Oriented Anatomy.* Lippincott Williams Wilkins, 2005.

[74] Thomas Netsch, Peter Rösch, Juergen Weese, Arianne van Muiswinkel, and Paul Desmedt. Grey value-based 3-D registration of functional MRI time-series: Comparison of interpolation order and similarity measure. In *Proceedings of SPIE (Vol. 3979) Medical Imaging 2000: Image Processing*, pages 1148–1159, 2000.

[75] Linda W. Nunes, Mitchell D. Schnall, Susan G Orel, Mary G. Hochman, Curtis P. Langlotz, Carol A. Reynolds, and Michael H. Torosian. Breast MR imaging: interpretation model. *Radiology*, 202:833–841, 1997.

[76] Nancy A. Obuchowski. Receiver operating characteristic curves and their use in radiology. *Radiology*, 229:3–8, 2003.

[77] Institute of Medicine. *Mammography and Beyond: Developing Technologies for the Early Detection of Breast Cancer.* National Academies Press, 2001. http://www.nap.edu/books/0309072832/html/.

[78] Institute of Medicine. *Saving women's lives: strategies for improving breast cancer detection and diagnosis*. National Academies Press, June 2004. http://www.iom.edu/report.asp?id=20721.

[79] American College of Radiology. *American College of Radiology (ACR) Breast Imaging Reporting and Data System Atlas (BI-RADS ®Atlas)–MRI*. Reston, VA, first edition, 2003.

[80] Susan G. Orel and Mitchell D. Schnall. MR imaging of the breast for the detection, diagnosis, and staging of breast cancer. *Radiology*, 220:12–30, 2001.

[81] Nobuyuki Otsu. A threshold selection method from gray-level histogram. *IEEE Transactions on System, Man, and Cybernetics*, 9(1):62–66, 1979.

[82] Sophie Paquerault, Yulei Jiang, Robert M. Nishikawa, Robert A. Schmidt, Carl J. D'Orsi, Carl J. Vyborny, and Gillian M. Newstead. Automated selection of BI-RADS lesion descriptiors for reporting calcifications in mammograms. In *Proceedings of SPIE (Vol. 5032) Medical Imaging 2003: Image Processing*, pages 802–809, 2003.

[83] Jeong Mi Park, Edmund A. Franken, Megha Garg, Laurie L. Fajardo, and Loren T Niklason. Breast tomosynthesis: Present considerations and future applications. *RadioGraphics*, 27:S231–S240, 2007.

[84] E. Parzen. On estimation of a probability density function and mode. *Annulal Mathematical Stastitics*, 33:1065–1067, 1962.

[85] Stephanie K. Patterson. New technologies in breast imaging. *Journal of the National Comprehensive Cancer Network*, 1(2):272–278, April 2003.

[86] Styliani Petroudi, Georgios Ketsetzis, and Michael Brady. Multi-vector segmentation of breast MR image via hidden markov random fields. In *Proceedings of the 8th WSEAS CSCC*, 2004.

[87] Josien P. W. Pluim, J. B. Antoine Maintz, and Max A. Viergever. Mutual-information-based registration of medical images: A survey. *IEEE Transactions on Medical Imaging*, 22(8):986–1004, August 2003.

[88] Alain Rakotomamonjy. Variable selection using SVM-based criteria. *Journal of Machine Learning Research*, 3:1357–1370, 2003.

[89] S. C. Rankin. MRI of the breast. *The British Journal of Radiology*, 73:806–818, August 2000.

[90] Charles L. Robertson. A private breast imaging practice: Medical audit of 25,788 screening and 1,077 diagnostic examinations. *Radiology*, 187:75–79, 1993.

[91] Daniel Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes. Non-rigid registration using free-form deformations: Application to breast MR images. *IEEE Transactions on Medical Imaging*, 18(8):712–721, 1999.

[92] Mehul P. Sampat, Alan C. Bovik, and Mia K. Markey. Classification of mammographic lesions into BI-RADS shape categories using the beamlet transform. In *Proceedings of SPIE Medical Imaging 2005: Image Processing*, 2005.

[93] Mehul P. Sampat, Mia K. Markey, and Alan C. Bovik. Computer-aided detection and diagnosis in mammography. In Alan C. Bovik, editor, *Handbook of Image and Video Processing*, pages 1195–1217. Academic Press, 2005.

[94] Mitchell D. Schnall, Jeffrey Blume, David A. Bluemke, Gia A. DeAngelis, Nanette DeBruhl, Steven Harms, Sylvia H. Heywang-Köbrunner, Nola Hylton, Christiane K. Kuhl, Etta D. Pisano, Petrina Causer, Stuart J. Schnitt, David Thickman, Carol B. Stelling, Paul T. Weatherall, Constance Lehman, and Constantine A. Gatsonis. Diagnostic architectural and dynamic features at breast MR imaging: Multicenter study. *Radiology*, 238:42–53, 2006.

[95] Mitchell D. Schnall and Debra Ikeda. Lesion diagnosis working group report of the international working group on breast MRI. *Journal of Magnetic Resonance Imaging*, 10:982–990, 1999.

[96] Frank G. Shellock and Albert Spinazzi. MRI safety update 2008: Part I, MRI contrast agents and nephrogenic systemic fibrosis. *American Journal of Roentgenology*, 191:1–11, October 2008.

[97] Bernard W. Silverman. *Density estimation for statistics and data analysis*. Chapman and Hall, 1986.

[98] American Cancer Society. *Cancer Facts and Figures 2006*. 2006.

[99] Mark J. Stoutjesdijk, Carla Boetes, and Gerrit J. Jager. Magnetic resonance imaging and mammography in women with a hereditary risk of breast cancer. *Journal of National Cancer Institute*, 93:1095–1102, 2001.

[100] Botond K. Szabó, Peter Aspelin, and Maria Kristoffersen Wiberg. Neural network approach fo the segmentation and classification of dynamic magnetic resonance images of the breast: comparison with empiric and quantitative kinetic parameters. *Academic Radiology*, 11(12):1344–1354, 2004.

[101] Botond K. Szabó, Peter Aspelin, Maria Kristoffersen Wiberg, and Beata Boné. Dynamic MR imaging of the breast: Analysis of kinetic and morphologic diagnostic criteria. *Acta Radiologica*, 44:379–386, 2003.

[102] Georgia D. Tourassi, Mia K. Markey, Joseph Y. Lo, and Carey E. Floyd Jr. A neural network approach to breast cancer diagnosis as a constraint satisfaction problem. *Medical Physics*, 28(5):804–811, 2001.

[103] Jayaram K. Udupa, Vicki R. LaBlanc, Hilary Schmidt, Celina Imielinska, Punam K. Saha, George J. Grevera, Ying Zhuge, L. M. Currie, Pat Molholt, and Yinpeng Jin. Methodology for evaluating image-segmentation

algorithms. In *Proceedings of SPIE (Vol. 4684) Medical Imaging 2002: Image Processing*, pages 266–277, 2002.

[104] T. W. Vomweg, M. Buscema, H. U. Kauczor, A. Teifke, M. Intraligi, S. Terzi, C. P. Heussel, T. Achenbach, O. Rieker, D. Mayer, and M. Thelen. Improved artifical neural networks in prediction of malignancy of lesions in contrast-enhanced MR-mammography. *Medical Physics*, 13(9):2350–2359, 2003.

[105] Jue Wang, Bo Thiesson, Yingqing Xu, and Michael Cohen. Image and video segmentation by anisotropic kernel mean shift. In *European Conference on Computer Vision*, pages 238–249, 2004.

[106] Simon K. Warfield, Kelly H. Zou, and William M. Wells. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Transactions on Medical Imaging*, 23(7):903–921, 2004.

[107] P.T. Weatherall, Gregory F. Evans, Gregory J. Metzger, M. Hossein Saborrian, and A. Marilyn Leitch. MRI vs. histologic measurement of breast cancer following chemotherapy: Comparison with x-ray mammography and palpation. *Journal of Magnetic Resonance Imaging*, 13:868–875, 2001.

[108] U. Wedegärtner, U. Bick, K. Wörtler, E. Rummeny, and G. Bongartz. Differentiation between benign and malignant findings of MR-mammography:

usefulness of morphological criteria. *European Radiology*, 11:1645–1650, 2001.

[109] Gary J. Whitman, Decland G. Sheppard, Michael J. Phelps, and Bianca N. Gonzales. Breast cancer staging. *Seminar in Roentgenology*, 41(2):91–104, 2006.

[110] Gert Wollny and Frithjof Kruggel. Computation cost of non-rigid registration algorithms based fluid dynamics. *IEEE Transactions on Medical Imaging*, 21(8):946–952, August 2002.

[111] Dafang Wu and Sanjiv Sam Gambhir. Positron emission tomography in diagnosis and management of invasive breast cancer: Current status and future persepctives. *Clinical Breast Cancer*, 4:S55–S63, 2003.

[112] Qiu Wu and Chandrajit Bajaj. B-spline representation for volume reconstruction. Technical report, The University of Texas at Austin, 2004.

[113] Qiu Wu and Mia K. Markey. Computer-aided diagnosis of breast cancer on MR imaging. In Jasjit S. Suri and Rangaraj M. Rangayyan, editors, *Recent Advances in Breast Imaging, Mammography, and Computer-Aided Diagnosis of Breast Cancer*, pages 743–766. SPIE Press, 2006.

[114] Qiu Wu, Marcos Salganicoff, Arun Krishnan, Donald S. Fussell, and Mia K. Markey. Interactive lesion segmentation on dynamic contrast enhanced breast MR using a Markov model. In *Proceedings of SPIE*

(Vol. 6144) Medical Imaging 2006: Image Processing, pages 1487–1494, 2006.

[115] Chenyang Xu and Jerry L. Prince. Snakes, shapes, and gradient vector flow. *IEEE Transactions on Image Processing*, 7(3):359–369, March 1998.

[116] Wei Tse Yang, Selin Carkaci, Lingyun Chen, Chao-Jen Lai, Aysegul Sahin, Gary J. Whitman, and Chris C. Shaw. Dedicated cone-beam breast CT: Feasibility study with surgical mastectomy specimens. *American Journal of Roentgenology*, 189:1312–1315, 2008.

[117] Barbara Zangheri, Cristina Messa, Maria Picchio, Luigi Gianolli, Claudio Landoni, and Ferruccio Fazio. PET/CT and breast cancer. *European Journal of Nuclear Medicine and Molecular Imaging*, 31:S136–S142, 2004.

Index

Vita

Qiu Wu was born in Anhui Province, China. He received the BS degree in Computer Science from Hefei University of Technology and Master of Engineering degree in Electrical Engineering (Signal Processing Track) from Southeast University, China. He had internship with Siemens Corporate Research (Princeton, NJ) and with Siemens Medical Solutions USA (Malvern, PA), in the summer of 2000 and summer of 2004, respectively. He was elected president(2003–2004) of the Graduate Engineering Council at The University of Texas at Austin. He served on Technical Program Committee of 2006 IEEE Region 5 annual conference. He is a recipient of a 2008 (Inaugural) Engineering Graduate Student Leadership Award from the Cockrell School of Engineering and a 2006 New Entrant Stipend Award from International Society of Magnetic Resonance in Medicine.

Permanent address: 7205 Hart Ln.
Austin, Texas 78731

This dissertation was typeset with LaTeX[†] by author.

[†]LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.